

# A Fixed-Point Approach to Equilibrium Pricing in Differentiated Product Markets

by

**William Ross Morrow**

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Mechanical Engineering)  
in The University of Michigan  
2008

Doctoral Committee:

Associate Professor Steven J. Skerlos, Chair

Professor Fred M. Feinberg

Professor Panos Y. Papalambros

Associate Professor Divakar Viswanath

© William Ross Morrow

---

All Rights Reserved

2008

# Acknowledgments

Besides my committee, I would like to acknowledge my wife, Erin MacDonald, my parents, Bob and Susan Morrow, and my many friends and colleagues for their consistent support.

# Table of Contents

|   |     |
|---|-----|
| <b>Acknowledgments</b> . . . . .  | ii  |
| <b>List of Tables</b> . . . . .   | vi  |
| <b>List of Figures</b> . . . . .  | vii |
| <b>List of Appendices</b> . . . . .   | ix  |
| <b>Abstract</b> . . . . .   | x   |
| <b>Chapter 1 Introduction</b> . . . . .   | 1   |
| 1.1 Analysis of Regulatory Policy . . . . .   | 2   |
| 1.1.1 Technology/Cost Analysis . . . . .  | 3   |
| 1.1.2 The Importance of Economic Behavior . . . . .   | 5   |
| 1.1.3 Heterogeneous Firms and Vehicle Model Detail . . . . .  | 10  |
| 1.2 Game-Theoretic Models of Product Design and Pricing . . . . .   | 12  |
| 1.2.1 Bertrand-Nash Equilibrium Prices . . . . .  | 12  |
| 1.2.2 Single-Stage Equilibrium Design and Pricing . . . . .   | 13  |
| 1.2.3 Multi-Stage Equilibrium Design and Pricing . . . . .  | 14  |
| 1.3 The Existing Understanding of Equilibrium Prices . . . . .  | 16  |
| 1.3.1 Theoretical Understanding . . . . .   | 17  |
| 1.3.2 Computation of Equilibrium Prices . . . . .   | 19  |
| 1.3.3 Empirical Validity . . . . .  | 22  |
| 1.4 Contributions Made in this Dissertation . . . . .   | 23  |
| <b>Chapter 2 A General Framework for Bertrand Competition Under<br/>an Arbitrary Random Utility Model</b> . . . . . | 28  |
| 2.1 Random Utility Models and Demand for Products . . . . .   | 29  |
| 2.2 Firms, Product Portfolios, Costs, and Profits . . . . .   | 31  |
| 2.3 Local Equilibrium and the Simultaneous Stationarity Conditions . . . . .  | 33  |
| <b>Chapter 3 An Analysis of Unregulated Equilibrium Prices Under<br/>Logit Models</b> . . . . .                     | 35  |
| 3.1 Logit Price Equilibrium . . . . .   | 35  |

|   |   |           |
|---|---|-----------|
| 3.1.1   | Systematic Utility Specifications . . . . .                                     | 38        |
| 3.1.2   | Logit Choice Probabilities and Profits . . . . .                                | 39        |
| 3.1.3   | Bounded and Vanishing Logit Profits . . . . .                                   | 41        |
| 3.1.4   | Fixed-Point Characterizations of Price Equilibrium . . . . .                    | 43        |
| 3.1.5   | Sufficiency of the Fixed-Point Characterizations . . . . .                      | 45        |
| 3.1.6   | Existence of Fixed-Points . . . . .   | 47        |
| 3.2   | Structural Properties of Price Equilibrium . . . . .                            | 49        |
| 3.2.1   | Intra-Firm Structural Properties . . . . .                                      | 49        |
| 3.2.2   | An Inter-Firm Structural Property . . . . .                                     | 53        |
| 3.3   | Conclusions . . . . .   | 54        |
| <b>Chapter 4 Computation of Unregulated Equilibrium Prices Under Mixed Logit Models of Demand . . . . .</b> |   | <b>55</b> |
| 4.1   | Fixed-Point Equations for Equilibrium Prices Under Mixed Logit Models . . . . . | 56        |
| 4.1.1   | Systematic Utility Specifications . . . . .                                     | 56        |
| 4.1.2   | Mixed Logit Choice Probabilities and Their Derivatives . . . . .                | 57        |
| 4.1.3   | Fixed-Point Equations . . . . .   | 58        |
| 4.2   | Numerical Methods . . . . .   | 60        |
| 4.2.1   | Preliminary Considerations . . . . .  | 61        |
| 4.2.2   | $\zeta$ Fixed-Point Iteration . . . . .   | 62        |
| 4.2.3   | Some Comments on the $\eta$ Map . . . . .                                       | 62        |
| 4.2.4   | Newton's Method . . . . .   | 63        |
| 4.3   | A Numerical Example: the 2005 New Vehicle Market . . . . .                      | 66        |
| 4.3.1   | Vehicle Data . . . . .  | 66        |
| 4.3.2   | Two Demand Models . . . . .   | 68        |
| 4.4   | Computing Equilibrium Prices with the Fixed-Point Iteration . . . . .           | 69        |
| 4.5   | Method Comparison . . . . .   | 72        |
| 4.6   | Conclusions . . . . .   | 80        |
| <b>Chapter 5 Regulated Equilibrium Prices Under Mixed Logit Models . . . . .</b>                            |   | <b>82</b> |
| 5.1   | Additions to the Framework . . . . .  | 82        |
| 5.1.1   | Regulatory Costs and Heuristic Expected Profits . . . . .                       | 82        |
| 5.2   | Several Regulatory Policies . . . . .   | 84        |
| 5.2.1   | Fuel Taxes. . . . .   | 84        |
| 5.2.2   | CO <sub>2</sub> taxes on firms. . . . .   | 85        |
| 5.2.3   | Corporate Average Fuel Economy (CAFE) Standards. . . . .                        | 86        |
| 5.2.4   | Corporate Average Emissions (CAE) Standards. . . . .                            | 88        |
| 5.2.5   | Feebates. . . . .   | 89        |
| 5.3   | Stationarity . . . . .  | 90        |
| 5.3.1   | Smooth Stationarity . . . . .   | 90        |
| 5.3.2   | Standards and Non-Smooth Stationarity . . . . .                                 | 92        |
| 5.3.3   | Simultaneous Stationarity . . . . .   | 95        |
| 5.4   | A Hybrid Fixed-Point Iteration . . . . .  | 95        |
| 5.5   | Conclusions . . . . .   | 99        |

|                     |   |     |
|---------------------|---|-----|
| <b>Chapter 6</b>    | <b>Conclusions and Future Work</b>        | 100 |
| 6.1                 | Criticisms of this Research               | 101 |
| 6.2                 | Future Work                               | 102 |
| 6.2.1               | Price Equilibrium                         | 102 |
| 6.2.2               | Single- and Multi-Stage Design Games      | 106 |
| 6.2.3               | A Roadmap for Policy Analysis             | 108 |
| <b>Appendices</b>   |   | 110 |
| B.1                 | Demonstrations of the Basic Definitions   | 113 |
| B.2                 | Examples of the $\zeta$ map               | 114 |
| C.1                 | Our Fixed-Point Iteration                 | 116 |
| C.2                 | Newton's Method                           | 117 |
| C.2.1               | Computing Jacobian Matrices               | 117 |
| C.2.2               | Quasi-Newton Methods and the BFGS Update. | 120 |
| C.2.3               | Inexact Newton Methods                    | 121 |
| D.1                 | Proofs for Chapter 3                      | 134 |
| D.2                 | Proofs for Chapter 4                      | 140 |
| D.3                 | Proofs for Chapter 5                      | 145 |
| <b>Bibliography</b> |   | 152 |

# List of Tables

|              |  |    |
|--------------|--|----|
| <b>Table</b> |  |    |
| 1.1          | One counterfactual experiment result from Beresteanu and Li (2008).  | 22 |
| 3.1          | Assumptions required for important profit maximizations results. . . .   | 37 |
| 3.2          | Assumptions required for important equilibrium results. . . . .  | 37 |
| 4.1          | Make statistics for the 2005 dataset. . . . .  | 67 |
| 4.2          | “Worst-case” observed computation times for various sample set sizes.  | 69 |
| 4.3          | Summary of results for ten perturbation trials under the Boyd and Mellman (1980) model for a fixed set of 1,000 samples. . . . . | 76 |
| 4.4          | Summary of results for ten perturbation trials under the Berry et al. (1995) model for a fixed set of 1,000 samples. . . . .     | 78 |
| 4.5          | Results of price equilibrium computations starting at unit costs under both demand models for ten 1,000-sample sets. . . . .     | 79 |
| 5.1          | Parameters proposed by NHTSA for attribute-based CAFE standards in 2011-2015. . . . .  | 88 |

# List of Figures

## Figure

|      |   |    |
|------|---|----|
| 4.1  | Transaction versus equilibrium prices for 993 vehicle models for a single 1,000,000 sample set under the Boyd and Mellman (1980) model. . . .   | 70 |
| 4.2  | Transaction versus equilibrium prices for 993 vehicle models for a single 1,000,000 sample set under the Berry et al. (1995) model. . . . .   | 70 |
| 4.3  | Change in $\ (\tilde{\nabla}\hat{\pi})(\mathbf{p})\ _{\infty}$ for 10 random initial conditions and a fixed sample set ( $S = 100,000$ ) under the Boyd and Mellman (1980) model.                   | 71 |
| 4.4  | Change in $\ (\tilde{\nabla}\hat{\pi})(\mathbf{p})\ _{\infty}$ for 10 random initial conditions and a fixed sample set ( $S = 100,000$ ) under the Berry et al. (1995) model. . . . .               | 71 |
| 4.5  | Standard deviation in computed equilibrium prices over 10 random initial conditions for several sample set sizes (given in the numeric labels) under the Boyd and Mellman (1980) model. . . . .     | 71 |
| 4.6  | Standard deviation in computed equilibrium prices over 10 random initial conditions for several sample set sizes (given in the numeric labels) under the Berry et al. (1995) model. . . . .         | 71 |
| 4.7  | Cumulative distribution of standard deviation in computed equilibrium prices over 10-20 sample sets of several sizes (given in the numeric labels) under the Boyd and Mellman (1980) model. . . . . | 72 |
| 4.8  | Cumulative distribution of standard deviation in computed equilibrium prices over 10-20 sample sets of several sizes (given in the numeric labels) under the Berry et al. (1995) model. . . . .     | 72 |
| 4.9  | Typical convergence curves for perturbation trials under the Boyd and Mellman (1980) model. . . . .   | 75 |
| 4.10 | Typical convergence curves for perturbation trials under the Boyd and Mellman (1980) model. . . . .   | 77 |
| 4.11 | Convergence curves for the fixed-point iteration and variants of Newton's method started at unit costs for a fixed 1,000 sample set under the Boyd and Mellman (1980) model. . . . .                | 79 |
| 4.12 | Convergence curves for the fixed-point iteration and variants of Newton's method started at unit costs for a fixed 1,000 sample set under the Berry et al. (1995) model. . . . .                    | 80 |
| 5.1  | A hybrid fixed-point approach for standard-regulated Bertrand competition. . . . .  | 97 |



|     |  |     |
|-----|--|-----|
| 5.2 | Convergence curves for several regulated fixed-point iterations under the Berry et al. (1995) model. . . . .                           | 98  |
| C.1 | Relative error in computed solutions to the Newton system and its preconditioned form using GMRES under the Berry et al. (1995) model. | 126 |

# List of Appendices

## Appendix

|   |   |     |
|---|---|-----|
| A | Mathematical Notation . . . . .                       | 111 |
| B | Examples for the Logit Model . . . . .                | 113 |
| C | An Extended Discussion of Numerical Methods . . . . . | 116 |
| D | Proofs . . . . .                                      | 134 |

# Abstract

The use of passenger cars in the U.S. is widely recognized as a primary source of greenhouse gas emissions. Among other concerns related to transportation energy usage, these emissions have motivated policy makers to reform regulatory policies that impact the design and pricing behavior of automotive firms. However many prominent analyses supporting these reforms neglect aspects of imperfect competition that many economists have come to recognize in the automotive industry, particularly Bertrand competition. Understanding the impacts of alternative regulatory policies accounting for imperfectly competitive behavior may improve policy makers' decisions regarding regulatory policy, as well as firms' abilities to profitably respond to the policies chosen. Examining and exploiting this potential first requires further development of the theory of Bertrand competition.

This dissertation makes three contributions the application of Bertrand competition in regulated differentiated product markets. We focus on the classes of Logit and Mixed Logit Discrete Choice Random Utility Models of consumer demand and avoid assumptions on the number or type of products offered by different firms. Our first contribution is a proof of the existence of unregulated equilibrium prices for Logit models using a new fixed-point equation equivalent to the first-order necessary condition for equilibrium, minimal assumptions on the utility specification, and mathematical tools from differential topology. This fixed-point equation is then generalized to the class of Mixed Logit models, one of the most flexible and popular empirical forms for representing consumer demand, under a weak hypothesis on the utility specification and mixing distribution. Several numerical approaches based on fixed-point characterizations of stationarity are demonstrated to be efficient and reliable methods for the computation of unregulated equilibrium prices in large-scale and complex differentiated product markets. Finally, we further extend this fixed-point approach to regulated equilibrium pricing problems with regulatory policy forms inspired by those considered for the U.S. automotive industry. Modified fixed-point iterations are derived for a number of regulatory policies with differentiable regulatory costs. A

hybrid fixed-point approach is developed for the computation of regulated equilibrium prices under standard-based policies with non-differentiable regulatory costs, such as the Corporate Average Fuel Economy standards currently active in the U.S.

# Chapter 1

## Introduction

Understanding the prices firms choose to set for the products they offer to market has long been a central task of economics. Of course, the most basic theory is the equality of supply and demand. In this theory, firms exercise no control over the prices they receive for their products or the quantities they offer to the market. Bertrand first introduced the idea that firms selfishly select those prices that maximize their own profits given the prices their competitors have chosen, committing to producing enough units of every product offered to meet whatever demand arises under these prices (Shapiro, 1989; Baye and Kovenock, 2008). Equilibrium prices in Bertrand competition are not merely prices that make markets clear, because firms can control production to ensure that at any prices. Rather, equilibrium prices are prices that simultaneously maximize profits for all firms in the industry. This model was originally a response to Cournot competition, which proposed that firms strategically control the quantities of the product they have in hand, letting prices be determined so that all products are sold. These models have been generalized and widely applied through the field of Game Theory. See, for example, Nash (1950, 1951); Luce and Raiffa (1957); Fudenberg and Tirole (1991); Myerson (1997).

Many authors consider the automotive market to be a classical example of an imperfectly competitive oligopoly, and subsequently characterize it with game-theoretic models such as Bertrand or Cournot competition (Bresnahan, 1987; Berry et al., 1995; Goldberg, 1995, 1998; Sudhir, 2001; Petrin, 2002; CBO, 2003; Berry et al., 2004; Austin and Dinan, 2005; Bento et al., 2005; Jacobsen, 2006; Beresteanu and Li, 2008). This market also plays a major role in the U.S.'s resource consumption and greenhouse gas emission problems. Motivated by the increasing intensity and public recognition of these problems, debate over how the automotive industry should be regulated has intensified over the past decade. Much of this debate has been shaped by quantitative

analyses of vehicle technologies that can improve fuel economy and emissions intensity (NRC, 2002; NESCCAF, 2004; EIA, 2003). However, as we explain below, these analyses neglect important aspects of economic behavior that game-theoretic models of design and pricing intend to capture. While there are game-theoretic analyses that study the impacts of regulatory policies, there remain both theoretical and pragmatic gaps to widespread implementation of game-theoretic models in the analysis of regulatory policy for differentiated product markets. This dissertation contributes to the construction and use of game-theoretic models of design and pricing by developing the underlying theory of equilibrium pricing and investigating methods for solving large-scale equilibrium pricing problems.

This introductory chapter motivates the contributions contained in Chapters 3-5. We begin in Section 1.1 with a discussion of influential quantitative analyses of regulatory policy options for the automotive industry. Our intention here is to justify the application of game-theoretic models of firm behavior in the analysis of regulated differentiated product markets. Next, in Section 1.2 we provide a formal description of game-theoretic models of product design and pricing. Here our emphasis lies on establishing the importance of equilibrium pricing in the game-theoretic analysis of regulated differentiated product markets. In Section 1.3 we motivate our specific contributions to equilibrium pricing with a discussion of the existing understanding of equilibrium prices for the largest class of empirically-relevant models of consumer demand. Finally Section 1.4 introduces our contributions in greater detail and outlines the remainder of this thesis.

## 1.1 Analysis of Regulatory Policy

The product offerings in the new car market are currently subject to well-known regulatory policies that are an important component of managing the growing fuel consumption and greenhouse gas emissions from personal vehicles. Congress debated increasing the Corporate Average Fuel Economy (CAFE) Standards (NHTSA, 2008b) as recently as last year, ultimately raising the standard levels<sup>1</sup> and increasing the size of vehicles regulated under the policy (NHTSA, 2008a). There are other approaches being debated for regulating the automotive industry. In 2005, California proposed to regulate CO<sub>2</sub> emissions (and other greenhouse gases) from vehicles sold in that state directly using a Corporate Average Emissions (CAE) standard (CARB, 2008).

---

<sup>1</sup>Loosely speaking, the standards were raised to an estimated 35 miles per gallon, although there is no strict standard under the new policy. This is explained in detail below.

Perceiving this to be too closely related to a fuel economy standard, which California has no legal authority to impose, this regulation was recently ruled down (Maynard, 2007). Cap and trade and carbon taxes have also been proposed (Dingell, 2007; Stavins, 2007).

It is fundamentally important that policy makers understand the potential consequences of alternative policy decisions when reviewing policy options. Firms operating in regulated markets also need this understanding in order to undertake strategic planning and decision making. Quantitative modeling plays a central role in both developing this understanding and justifying particular decisions. In this section we provide a brief overview of several quantitative analyses that have influenced the recent automotive policy debate. Our intention is to convince the reader that game-theoretic models of regulation product design and pricing should be a component of the analysis of regulatory policy in differentiated product markets like automotive.

### 1.1.1 Technology/Cost Analysis

In a review of 20 years of analyses focusing on the impact of vehicle technologies on fuel economy, Greene and DeCicco (2000) define the core of Technology/Cost Analysis, the prevailing policy analysis paradigm, as follows:

“Technology/Cost Analysis is a method for estimating the potential of automotive fuel economy improvement by enumerating specific technologies, estimating their cumulative impact on fuel economy, and adding up their cost.” (Greene and DeCicco, 2000, pg. 480)

Most often such studies consider aggregate or “average” vehicles, rather than the thousands of specific vehicle models actually offered by firms, and account for the impact of technology adoption on fuel economy and cost using incremental percentage improvements. Marketshares of particular technologies within an aggregate vehicle class are projected and combined into projections of the impact of technological diffusion on vehicle fuel economy and cost using simple rules for estimating the combined impact of multiple technologies. Here we briefly review three prominent applications of these principles.

The U.S. National Research Council (NRC) employed the Technology/Cost paradigm in their prominent 2001/2002 review of the CAFE standards (NRC, 2002), a study that continued to be influential in last year’s Congressional debates on CAFE reform (Levin, 2007b) and has shaped the new structure of the CAFE standards (NHTSA, 2006, 2008a). In this study, twelve aggregate vehicle types were defined.

Incremental vehicle technologies were identified and their cumulative potential impact on fuel economy projected. Incremental costs for these technologies were estimated by reviewing literature and interviewing suppliers.

In 2004 the Northeast States Center for a Clean Air Future (NESCCAF) released a report detailing incremental vehicle technology *packages* that could reduce vehicle greenhouse gas emissions. California’s Air Resources Board relied on this study heavily when designing their proposed CAE standard (CARB, 2008). The authors of the NESCCAF report defined five vehicle types and, instead of applying percentage improvements, applied the vehicle simulation software CRUISE (AVL, 2008b) to assess incremental emissions intensity reductions of feasible vehicle technology packages selected for each of these vehicle types. This simulation-based assessment approach was intended to capture technical synergies and conflicts that could impact performance improvements and might have been neglected in the NRC study. Incremental costs for adopting these technology packages were estimated by interviewing suppliers. The potential impact of these technologies on other vehicle characteristics (e.g., size, weight, power, and speed) was largely ignored. In fact, vehicle acceleration was purposely held constant.

Neither the NRC nor the NESCCAF report aims to project a specific fuel economy or emissions intensity level will be achieved under a particular regulatory policy structure. These studies focus almost entirely on technology, aiming only to make claims to the technical feasibility of a certain level of improvement in fuel consumption or emissions intensity. In the end, both studies conclude that there are large feasible improvements in fuel economy and emissions intensity to be gained through existing vehicle technologies, and that the additional costs of these technologies are likely to be balanced by future fuel savings. While this is an appropriate use of Technology/Cost Analysis, policy makers choosing regulatory policy options to achieve reduced fuel consumptions and greenhouse gas emissions, as well as the firms being regulated, must understand the systemic economic consequences of various vehicle technologies and regulatory policies.

The National Energy Modeling System (NEMS), a computable general equilibrium model for the entire U.S. economy focused on energy demand and supply (EIA, 2003, 2007), is aimed at this ultimate need for of integrated technological/economic analysis. The NEMS projects energy demand in the transportation sector using a mix of Technology/Cost and economic principles and has been repeatedly used to advise Congress on issues of transportation policy (EIA, 2008b). Only two aggregate vehicle types are considered: a car and a light truck. Energy demand from transportation depends on



the diffusion of specific fuel economy technologies. The specific technologies considered can be found in [EIA \(2008a\)](#). Technology marketshares for four aggregate vehicles – a domestic car and truck and a foreign car and truck – are projected based on cost effectiveness calculations that relate the “value of fuel savings” (and regulatory pressure for increased fuel economy) and the “value of performance” gained or lost by adopting the technology to the technology cost ([EIA, 2007](#), pg. 17-23).<sup>2</sup> We are not aware of any theoretical or empirical justification for the specific formulae applied provided in the model documentation. The NEMS also assumes that increased technology costs are passed on directly to consumers through vehicle prices, and that regulatory policy (i.e., fuel taxes and CAFE) have no direct effect on vehicle prices ([EIA, 2007](#), pg. 30).

## 1.1.2 The Importance of Economic Behavior

### Pricing

Increases in the costs of doing business aren’t always transmitted directly to consumers as NEMS presupposes. The airline industry has recently provided a convenient example of this. Due to the recent spike in world oil prices, carriers are facing drastic increases in the costs of doing business and are correspondingly raising prices. USA Today ([USAToday, 2008](#)) recently reported that summer airline fares are up 18% since last summer, and only on the most popular routes. However, fuel costs have risen by “50% or more” ([USAToday, 2008](#)). The International Air Travel Association has a much stronger estimate, claiming that jet fuel prices have essentially doubled ([IATA, 2008](#)). It is unlikely that unobserved cost-cutting can account for this large a gap between these increased costs and the corresponding price hikes. Rather, it is more likely that carriers perceive that, in the presence of their competitors, they cannot simply pass on all these costs to their customers without unacceptably negative impacts on profitability. Transliterating this observation to the automotive industry suggests that firms will absorb some of the costs of technology adoption, whether this is fueled by regulation or by changes in the marketplace.

Strategic pricing has generally become an important model of the automotive industry. While the nature of strategic interaction in the automotive market is not yet settled, a consistent conclusion appears to be that firms set prices to maximize their own profits. Particularly, there appears to be a strong consensus that the automotive

---

<sup>2</sup>Ad hoc corrections to marketshares and fuel economy impacts are made based on feasibility and synergy considerations.

market follows Bertrand competition within the “New Empirical Industrial Organization” (NEIO) literature; see, e.g., [Bresnahan \(1987\)](#); [Berry et al. \(1995\)](#); [Verboven \(1996\)](#); [Goldberg \(1995, 1998\)](#); [Fershtman et al. \(1999\)](#); [Sudhir \(2001\)](#); [Petrin \(2002\)](#); [Verboven \(2002\)](#); [Berry et al. \(2004\)](#); [Bento et al. \(2005\)](#); [Jacobsen \(2006\)](#); [Beresteanu and Li \(2008\)](#). Particularly, [Verboven \(1996\)](#) claims that

“The assumption of price-setting behavior in the car market is common and consistent with industry wisdom, see, e.g., the discussion of pricing practices by the U.K. Monopolies and Mergers Commission Report (1992).” ([Verboven, 1996](#), pg. 247)

Work from this area estimates unit costs and consumer demand, typically using Logit, Generalized Extreme Value, and Mixed Logit Random Utility Models (RUMs) ([McFadden, 1981](#); [Louvierre et al., 2000](#); [Train, 2003](#)). With these models, the impact of historical events or future changes can be evaluated using *counterfactual experiments*, in which the impact of changes to the economic environment relative to the environment-at-estimation are assessed. Such studies have provided insight into the welfare impacts of regulatory policy changes ([Goldberg, 1995](#); [Verboven, 1996](#); [Goldberg, 1998](#); [Fershtman et al., 1999](#); [Bento et al., 2005](#); [Beresteanu and Li, 2008](#)), corporate mergers ([Nevo, 2000a](#)), innovative products ([Petrin, 2002](#)), and other product portfolio changes ([Berry et al., 2004](#)).<sup>3</sup>

Generally speaking, regulatory policies have an impact on product prices in imperfectly competitive markets even prior to motivating the design of more efficient, and often more expensive, products and services. This is true even for regulatory policies directed at non-price characteristics of these products or do not even explicitly regulate the firm at all. For instance, economists have acknowledged the potential for the CAFE standards to influence pricing behavior in a specific way, a phenomenon called *mix-shifting*: in the “short run” firms should have the incentive to *raise* the prices of their *least* fuel efficient vehicles and *lower* the prices of their *most* fuel efficient vehicles in order to support movement of their Corporate Average Fuel Economy toward the standard level ([Goldberg, 1998](#); [Kleit, 1990, 2004](#); [Austin and Dinan, 2005](#); [Jacobsen, 2006](#)).

Out of the many models of the automotive market that apply Bertrand competition, a number intend to study the influence of regulatory policies like tax structures ([Verboven, 1996](#); [Fershtman et al., 1999](#); [Bento et al., 2005](#); [Verboven, 2002](#); [Beresteanu and Li, 2008](#)), import quotas ([Goldberg, 1995](#)), and the CAFE standards ([Goldberg,](#)

---

<sup>3</sup>In the last study, full equilibrium price computations were not undertaken in conflict with the equilibrium pricing assumptions generating the model.

1998; Jacobsen, 2006). These studies implicitly claim that the potentially complex pricing behavior resulting from imperfect competition in differentiated product markets needs to be an aspect of the assessment of alternative regulatory policies for these markets. Even though gasoline taxes don't explicitly regulate firm behavior at all, empirical studies have generated evidence that local sales taxes influence vehicle prices (Doyle, 1997), that firms may exploit tax reductions by raising prices (Fershtman et al., 1999), the expected reductions in demand effects from an increase in fuel taxes can be halved when models include firms that price strategically in response to tax changes (Verboven, 2002), and that the benefits of tax policy changes depend on how the tax revenue is recycled (Bento et al., 2005). Beyond contradicting the NEMS assumption that regulatory policy has no impact on prices for cars and trucks, these results illustrate important, unintended effects that could dampen the efficacy of gasoline taxes as a policy option for reduced GHG emissions.

Counterfactual experiments on Bertrand competition models are becoming a primary tool in analyzing potential impacts regulatory policy decisions. Particularly, Goldberg (1998) estimated a Bertrand competition model of the U.S. car market under CAFE and compared the effect of increases in CAFE to increases in gasoline taxes.<sup>4</sup> More recently, Bento et al. (2005) analysis of the consumer welfare impacts of increased gasoline taxes under different tax revenue recycling schemes applies a Bertrand competition model with Mixed Logit demand, assuming that all firms satisfy the CAFE standards. They simulate the impact of modest increases in gasoline taxes of 10, 30, and 50 cents per gallon, concluding that income-based recycling is highly regressive,<sup>5</sup> resulting in significant heterogeneity in welfare impacts across income groups and other demographic variables compared to tax-based recycling. Echoing Kleit's observation that not all firms respect the CAFE standards (Kleit, 1990, 2004), Jacobsen (2006) extends this modeling paradigm to study the impact of changes to the CAFE standards including firms that violate the standards. Most recently, Beresteanu and Li (2008) study hybrid vehicle purchasing during 2000-2006 using an empirical Bertrand competition model of the automotive market with a Mixed Logit demand model.<sup>6</sup> They report on a number of counterfactual experiments meant to study the effects of increased gasoline prices observed during the period, changes to the tax incentives for hybrid vehicles, and a flat rebate on hybrid vehicle sales.

---

<sup>4</sup>Unfortunately, Goldberg used an arithmetic average of fuel economy instead of the harmonic average required by CAFE legislation and thus her results are not necessarily reflective of the impacts of changes to the real CAFE standards.

<sup>5</sup>That is, the tax disproportionately impacts lower-income households.

<sup>6</sup>There referred to as a "random coefficients" model.

## Design

Recent Technology/Cost Analyses also neglect how the benefits of vehicle technologies are allocated to fuel economy versus other characteristics that are desired by consumers. [Greene and DeCicco \(2000\)](#) recognize this as a major obstacle for Technology/Cost Analysis. They give the simple example of electronic transmission control, which

“can be implemented to provide a small efficiency benefit along with improved shift smoothness and reliability. The same hardware can be reprogrammed to ... provide the greatest fuel efficiency benefit, but with a change in the feel of driving smoothness.” ([Greene and DeCicco, 2000](#), pg. 485)

The NRC also recognizes that most fuel economy improving technologies improve vehicle energy efficiency generally and are likely to be applied by firms to improve vehicle characteristics that are more strongly valued by consumers ([NRC, 2002](#)). GM is currently advertising their 2009 Cadillac Escalade hybrid as “intelligent indulgence” ([GM, 2008](#)), or efficiency without compromising aspects of luxury such as stereos and DVD players. While the hybrid powertrain will certainly improve the fuel economy of the Escalade, inclusion of such luxury features precludes achievement of maximal fuel economy gains.

The diversion of energy efficiency from fuel economy or emissions improvement can demonstrate how the complex linkage between technical aspects of product design and economic behavior of firms must be considered in effective assessments of regulatory policy. Suppose that regulatory policy incentivizes firms to product more efficient vehicles with declining or stagnant preferred performance characteristics. As a result, consumers are less likely to exchange their existing vehicles for new models.<sup>7</sup> Firms have an incentive to “relax” intensified price competition, arising from smaller markets, through product design ([Shaked and Sutton, 1982](#)). Diverting energy efficiency improvements from fuel economy to other performance characteristics of vehicles to attract consumers to new vehicle purchases is a probable mechanism for this relaxation, implying reduced fuel economy benefits (for the same costs).

Along these lines, [Greene and DeCicco](#) observe that some Technology/Cost Analyses include “markets for fuel economy,” intending to either improve the accuracy of fuel economy forecasts or to permit classical economic measurements of welfare impacts (to either consumers or firms, or both). The central input to such models are

---

<sup>7</sup>This may be changing with an evolving culture and the tripling of gasoline prices that has occurred in the intervening years and is likely to continue. Essentially, the “public” and “private” motivations may be aligning.

assumptions about firm’s decision processes and consumers’ preferences over different vehicle configurations. For example, “[f]uel economy supply curves ... reflect assumptions about producer behavior, such as cost minimization.” (Greene and DeCicco, 2000, pg. 489)

There exist several recent analyses meant to inform the debate surrounding the CAFE standards that include vehicle fuel economy choice, one aspect of design.<sup>8</sup> In order to compare CAFE standards to gasoline taxes, the Congressional Budget Office (CBO, 2003) and Austin and Dinan (2005) model firms choosing fuel economy and vehicle prices to maximize profits. As in Bento et al. (2005), all firms are assumed to respect the CAFE standards. The model framework is relatively simplistic. Unit costs as a function of fuel economy are estimated using NRC (2002) data on fuel economy improving technologies and their costs. Consumer demand is modeled using price elasticities of demand, rather than a RUM as in the wider literature concerning game-theoretic models of the automotive industry. Demand for fuel economy is connected to these price elasticities through discounted lifetime fuel costs.

In contrast to the assumptions adopted by the CBO (2003), Austin and Dinan (2005), and Bento et al. (2005), Kleit (1990, 2004) observes that U.S. automakers tend to unconditionally satisfy the CAFE standards while luxury European brands tend to treat the CAFE standards as a “mere tax” (Kleit, 2004, pg. 280).<sup>9</sup> U.S. firms contend<sup>10</sup>

“[w]ere they to violate the standards, ... they would be liable for civil damages in stockholder suits” (Kleit, 2004, pg. 280)

Kleit (1990) accommodates this behavior in the context of a simple Cournot competition model by imposing *ad hoc* constraints on firms optimal production problems. These constraints are used to estimate “shadow taxes” on firms constrained by the CAFE standards. Kleit (2004) goes further, including fuel economy decisions as strategic variables that firms constrained by the standard use to minimize total cost.<sup>11</sup> Kleit’s model frameworks are otherwise simplistic, considering only a few firms and

---

<sup>8</sup>While cost minimization and profit maximization are reasonable proposals, it is possible to make unrealistic assumptions about firm behavior. For example, a recent analysis of Feebates (Greene et al., 2005) assumes that manufacturers choose fuel economy to maximize *consumer’s* surplus, rather than maximizing their own profits or minimizing total costs, in conflict with much economic theory.

<sup>9</sup>Jacobsen (2006), following Kleit (2004), also applies this observation but does not consider firms that choose the design of their vehicles.

<sup>10</sup>The official record of CAFE standard violation is available from the National Highway Safety Administration (NHTSA) (NHTSA, 2008c). Kleit (1990, 2004) also notes that, as can be seen in this record, during the one period that prominent U.S. automakers were sure to violate the standards the policy was put on hold.

<sup>11</sup>It is not made clear how other firms choose fuel economy, if they do at all.

vehicles, applying linear supply and demand functions, and assuming that the vehicle market is perfectly competitive in that automotive firms have no control over the prices they set. [Austin and Dinan](#) openly critique this last aspect of [Kleit](#)'s modeling efforts.

While these attempts to integrate design decisions into economic models for policy analysis are a useful first step, they are not quite in line with the most prominent theoretical economic models. Particularly, design decisions are less flexible than prices, in the sense that design decisions have to be fixed well in advance of the point-of-sale, while prices are flexible up until this point, at least in principle. This fact might lead theoretical economists to model design and pricing with a multi-stage, as opposed to single-stage game.

The use of elasticity-based demand models is also somewhat less appropriate for a model in which firms are changing the new vehicle fleet. These demand models characterize demand for a fixed set of vehicles based on price alone, with parameters linked directly to the observed vehicle fleet. Thus, the appropriate demand parameters for a new product or even for an existing product with modified attributes are somewhat poorly defined. Furthermore, demand as a function of vehicle characteristics must be linked to the value of characteristic changes, as in [Kleit \(2004\)](#) and [Austin and Dinan \(2005\)](#), although this value is not consistently estimated with the parameters that govern the relationship between price and demand, if estimated at all. On the other hand, RUMs are designed to extract systematic relationships between various product characteristics, price, and demand for an observed vehicle fleet. Of course, one must assume that these systematic relationships do not change when the vehicle fleet changes when evaluating demand for new products or products with different attributes. Marketers and economists are currently very comfortable with this assumption.

### **1.1.3 Heterogeneous Firms and Vehicle Model Detail**

Another related aspect of real markets that is missing from the NRC, NESCCAF studies and the NEMS system is a characterization of the impact that different regulatory policies have on specific, real firms with heterogeneous product portfolios, capabilities, and costs. The role of such considerations in actually crafting policy in the U.S. should not be underestimated. For example, during the most recent round of CAFE debates the use of fleet-wide averages was successfully argued as discriminatory toward the American automakers who sell a disproportionate amount

of larger vehicles (NRC, 2002; Levin, 2007b,a). This has resulted in new firm-specific “attribute-based” standards tied to specific vehicle characteristics (NHTSA, 2008a, pg. 136).

As we have pointed out in this section, most Technology/Cost Analyses are reductive, studying only a few vehicle types and a few manufacturers. Some market-oriented studies (e.g., Kleit (1990, 2004); Bento et al. (2005); Jacobsen (2006)), are similarly reductive, studying only a few firms and vehicle types. However most econometric studies reject this reductivism, characterizing many firms and vehicles.

The National Highway Safety Administration (NHTSA), the federal government body to whom responsibility over CAFE has been delegated, has also rejected this reductive approach for its preliminary analysis of the new CAFE standards:

“For the model years covered by the current proposal, the light vehicle (passenger car and light truck) market forecast included more than 3,000 vehicle models, more than 400 specific engines, and nearly 400 specific transmissions. This level of detail in the representation of the vehicle market is vital to an accurate analysis of manufacturer- specific costs and the analysis of reformed CAFE standards, and is much greater than the level of detail used by many other models and analyses relevant to light vehicle fuel economy.” (NHTSA, 2008a, pg. 149)

While this level of detail is encouraging, NHTSA rejects profit-driven technology adoption and pricing in the analysis that has determined their regulatory levels. Specifically:

“The proposed standards were developed using a computer model (known as the ‘Volpe Model’) that, for any given model year, applies technologies to a manufacturers fleet until the manufacturer reaches compliance with the standard under consideration. The standards were tentatively set at levels such that, considering the seven largest manufacturers, the cost of the last technology application equaled the benefits of the improvement in fuel economy resulting from that application.” (NHTSA, 2008a, pg. 12)

This method is an attempt to determine “maximum feasible” regulatory levels that balance the benefits and costs of a particular regulatory level. However the lack of profit-driven technology adoption (i.e. design) and the corresponding imperfectly competitive pricing behavior cannot be expected to result in an accurate assessment of either the costs of the policy, to both consumers and firms, or the benefits of ultimate fuel consumption and emissions reduction under the proposed policy. It is these features of real markets that are captured by game-theoretic models of regulated vehicle design and pricing.

## 1.2 Game-Theoretic Models of Product Design and Pricing

In this section we formally introduce game-theoretic models of strategic design and pricing. These models are not the focus of the remainder of this dissertation. Rather through this discussion we hope to convince the reader that models of equilibrium pricing are a fundamental component of game-theoretic models of competitive design and pricing, thereby justifying the focus on equilibrium pricing that we do maintain.

To begin the development we first provide a brief formal definition of equilibrium prices. Our full model is described in Chapter 2.

### 1.2.1 Bertrand-Nash Equilibrium Prices

Suppose  $F \in \mathbb{N}$  firms offer  $J \in \mathbb{N}$  products to a consumer population. Each product  $j \in \mathbb{N}(J)$  is characterized by a vector of characteristics  $\mathbf{y}_j \in \mathcal{Y} \subset \mathbb{R}^K$  (for some  $K \in \mathbb{N}$ ) and a price  $p_j \in [0, \infty) = \mathbb{R}_+$ . Assume that some function  $\mathbf{q} : \mathcal{Y}^J \times \mathbb{R}_+^J \rightarrow \mathbb{R}_+^J$  gives the quantities of all products demanded by the population given any matrix of all product characteristics  $\mathbf{Y} \in \mathcal{Y}^J$  and vector of all prices  $\mathbf{p} \in \mathbb{R}_+^J$ . Each firm  $f \in \mathbb{N}(F)$  offers the  $J_f$  products indexed by the integers in  $\mathcal{J}_f \subset \mathbb{N}(J)$ , at unit costs  $\mathbf{c}_f^U \in \mathbb{R}_+^{J_f}$  and fixed cost  $c_f^F \in \mathbb{R}_+$ . For any firm  $f$  we denote its own prices by  $\mathbf{p}_f \in \mathbb{R}_+^{J_f}$  and the prices set by its competitors  $\mathbf{p}_{-f} \in \mathbb{R}_+^{J_{-f}}$ , where  $J_{-f} = \sum_{g \neq f} J_g$ .  $\mathbf{q}_f : \mathcal{Y}^J \times \mathbb{R}_+^J \rightarrow \mathbb{R}_+^{J_f}$  denotes the demand function for firm  $f$  products only.

With these definitions, every firm solves the following optimal pricing problem:

$$\begin{aligned}
 & \text{given } \mathbf{Y} \in \mathcal{Y}^J, \mathbf{p}_{-f} \in \mathbb{R}_+^{J_{-f}}, \mathbf{c}_f^U \in \mathbb{R}_+^{J_f}, c_f^F \in \mathbb{R}_+ \\
 & \text{maximize } \pi_f(\mathbf{Y}, \mathbf{p}) = \mathbf{q}_f(\mathbf{Y}, \mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f^U) - c_f^F \\
 & \text{with respect to } \mathbf{p}_f \in \mathbb{R}_+^{J_f} \\
 & \text{where } \mathbf{p} = (\mathbf{p}_f, \mathbf{p}_{-f})
 \end{aligned} \tag{1.1}$$

Note that the objective of this optimization problem depends on the choices of *all* firms, though firm  $f$  controls only a subset of the variables.

**Definition 1.2.1.**  $\mathbf{p}^*$  is a (Bertrand-Nash) equilibrium of the game formed by Problem (1.1) if  $\mathbf{p}_f^*$  solves Problem (1.1) for all  $f \in \mathbb{N}(F)$ . The set of all equilibrium prices, as a function of all product characteristics  $\mathbf{Y} \in \mathcal{Y}^J$  and all unit costs  $\mathbf{c}^U \in \mathbb{R}_+^J$  will be denoted by  $\mathfrak{E}(\mathbf{Y}, \mathbf{c}^U)$ .<sup>12</sup>

<sup>12</sup>As we will assume throughout this document, equilibrium prices do not depend on fixed costs.



Formally speaking, this is the definition of *pure strategy* equilibrium. Many of the early theoretical successes of Game Theory were induced by “convexifying” games through the introduction of mixed strategies, i.e. probability distributions over the set of actions. For example, it is only in mixed strategies that Nash’s seminal existence proof applies (Nash, 1950, 1951). However, applying mixed strategies to game-theoretic models of real firms may be dubious simply because “firms don’t play dice in the boardroom” (Choi et al., 1990).

### 1.2.2 Single-Stage Equilibrium Design and Pricing

As we have pointed out, firms operating in differentiated product markets also get to choose the physical characteristics of products as well as their prices in order to maximize profits or another chosen business objective. Beginning with Hotelling’s seminal oligopolistic model of two firms choosing where to locate stores in order to maximize profits (Hotelling, 1929), there have appeared many theoretical studies of firm behavior including market entry/exit and some qualitative form of product or service design in addition to pricing.<sup>13</sup>

The most direct way to extend the competitive pricing model to include product characteristics is to assume that firms choose both product characteristics and prices “simultaneously” with their competitors (see, e.g., Economides (1987)). The fuel economy and price choice models used by CBO (2003) and Austin and Dinan (2005) fall into this category. For another example, Michalek et al. (2004) provides a model of single-stage design and pricing that also includes regulation and “engineering content.”<sup>14</sup> Particularly, firms choose an engine type, engine size, and final drive ratio for a representative midsize vehicle. These three “technical characteristics” or “engineering variables” determine fuel economy and 0-60 acceleration, performance characteristics ultimately valued by consumers. The ADVISOR software (AVL, 2008a) (and interpolating splines) were used to map between engineering variables and performance characteristics. Unit and fixed costs were estimated as a function of the engineering variables from publicly available data. With a modified version of the Logit model estimated by Boyd and Mellman (1980) representing demand, firm profit functions are defined and coupled. Michalek et al. use this model to study the influence of alternative regulatory policies on the resultant equilibrium vehicle designs

---

<sup>13</sup>Hotelling’s purpose was to demonstrate a qualitative feature of some markets, the “Principle of Minimum Differentiation,” which states that firms strategically minimize the differentiation of their products. See also d’Aspremont et al. (1979). In real markets, firms both differentiate their products and make them similar to one another.

<sup>14</sup>See also Skerlos et al. (2005).

and prices, when designs and prices are chosen by all firms “at the same time.”

We can write a formal version of such a *single-stage* game as follows. Let  $\mathcal{X}$  be some set from which feasible values of engineering variables  $\mathbf{x}$  are taken, and let  $\tau : \mathcal{X} \rightarrow \mathcal{Y}$  be a “technology map” that defines performance characteristics in terms of these technical variables.<sup>15</sup> Suppose that the unit and fixed cost functions depend on the technical variables, as in Michalek et al. (2004), as opposed to the product characteristics.<sup>16</sup> Each firm solves the optimization problem:

$$\begin{aligned}
& \text{given} && \mathbf{Y}_{-f} \in \mathcal{Y}^{J-f}, \mathbf{p}_{-f} \in \mathbb{R}_+^{J-f} \\
& \text{maximize} && \pi_f(\mathbf{X}_f, \mathbf{Y}_{-f}, \mathbf{p}) = \mathbf{q}_f(\mathbf{Y}, \mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f^U(\mathbf{X}_f)) - c_f^F(\mathbf{X}_f) \\
& \text{with respect to} && (\mathbf{X}_f, \mathbf{p}_f) \in \mathcal{X}^{J_f} \times \mathbb{R}_+^{J_f} \\
& \text{where} && \mathbf{Y}_f = \tau(\mathbf{X}_f) \\
& && \mathbf{Y} = (\mathbf{Y}_f, \mathbf{Y}_{-f}) \\
& && \mathbf{p} = (\mathbf{p}_f, \mathbf{p}_{-f})
\end{aligned} \tag{1.2}$$

Pure strategy equilibria of such a game is defined exactly as in the case of Bertrand competition.

**Definition 1.2.2.**  $(\mathbf{X}^*, \mathbf{p}^*)$  is a (Nash) equilibrium of the game formed by Problem (1.2) if  $(\mathbf{X}_f^*, \mathbf{p}_f^*)$  solves Problem (1.2) for all  $f \in \mathbb{N}(F)$ .

### 1.2.3 Multi-Stage Equilibrium Design and Pricing

In reality, product design decisions are almost universally “less flexible” than pricing decisions. For example, vehicles design changes start at least 2 years in advance (Hill et al., 2007) while, in principle, market prices must be fixed only at the point of sale. Many economists, recognizing this as a common qualitative feature of many markets as far back as Hotelling’s original model, have applied *multi-stage* games to model competition among firms when the strategic variables firms control have varying degrees of flexibility; see, e.g., Shaked and Sutton (1982, 1990); Anderson and de Palma (1992a,b); Anderson et al. (1995); Verboven (1999); Anderson and de Palma (2001, 2006); Doraszelski and Draganska (2006). Economists have most often used this modeling paradigm to investigate sub-optimality of various market features –

<sup>15</sup>For simplicity, we assume  $\mathcal{X}$  and  $\tau$  are shared by all firms. This is a trivial assumption to relax, if desired. We also use  $\tau$  to refer to the vector function  $\tau : \mathcal{X}^N \rightarrow \mathcal{Y}^N$  defined by  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N) \mapsto (\tau(\mathbf{x}_1), \dots, \tau(\mathbf{x}_N)) = \tau(\mathbf{X})$  for any  $N \in \mathbb{N}$ .

<sup>16</sup>This is a conceptual (and computational) device, as firms could choose product characteristics rather than technical variables taking costs to be, for example,  $c^U(\mathbf{y}) = \inf \{c^U(\mathbf{x}) : \mathbf{x} \in \tau^{-1}(\mathbf{y})\}$ .

the number of firms, the “quality” of products, the “variety” of products – relative to certain socially-optimal states.

Loosely speaking, multi-stage games are predicated on the assumption that, when deciding on their “inflexible” strategic variables, firms recognize and account for the responses their competitors will have in their own “flexible” strategic variables to these choices. For example, a particular vehicle-designing firm understands that while its competitors may not be able to react to its vehicle design decisions with their own *vehicle design changes*, these competitors *can* react with *competitive pricing behavior* when the vehicles are ultimately offered to the public.

Formally speaking, multi-stage games are (*imperfect information*) *extensive form games* (Myerson, 1997). Analysts usually restrict attention to *subgame-perfect equilibrium* of multi-stage models of competing firms.<sup>17</sup> Loosely speaking, full (Nash) equilibrium in extensive form games is too broad a notion to be acceptable for realistic models of rational actors; see Myerson (1997).

Because defining subgame-perfect equilibrium in a general way is somewhat involved, we define it here only for the relatively simple example of two-stage design and pricing. If each firm assumes that, once all vehicle designs are fixed, pricing will occur in equilibrium, then each firm faces the following optimization problem:<sup>18</sup>

$$\begin{aligned}
& \text{given } \mathbf{Y}_{-f} \in \mathcal{Y}^{J-f}, \mathbf{c}_{-f}^U \in \mathbb{R}_+^{J-f} \\
& \text{maximize } \pi_f(\mathbf{X}_f, \mathbf{Y}_{-f}, \mathbf{p}) = \mathbf{q}_f(\mathbf{Y}, \mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f^U(\mathbf{X}_f)) - c_f^F(\mathbf{X}_f) \\
& \text{with respect to } (\mathbf{X}_f, \mathbf{p}) \in \mathcal{X}^{J_f} \times \mathbb{R}_+^J \\
& \text{subject to } \mathbf{p} \in \mathfrak{E}(\mathbf{Y}, \mathbf{c}^U) \tag{1.3} \\
& \text{where } \mathbf{Y}_f = \boldsymbol{\tau}_f(\mathbf{X}_f) \\
& \mathbf{Y} = (\mathbf{Y}_f, \mathbf{Y}_{-f}) \\
& \mathbf{c}^U = (\mathbf{c}_f^U(\mathbf{X}_f), \mathbf{c}_{-f}^U)
\end{aligned}$$

Note that because  $\mathfrak{E}$  need not be single-valued, solutions to Problem (1.3) are *pairs*  $(\mathbf{X}_f, \mathbf{p}) \in \mathcal{X}^{J_f} \times \mathbb{R}_+^J$ . We also remark that competitors *costs* are required to evaluate the equilibrium constraint in Problem (1.3). Subgame-perfect equilibria are defined as follows.

**Definition 1.2.3.** *A subgame-perfect equilibrium of the game formed by Problem (1.3)*

<sup>17</sup>Strictly speaking, most models, including ours, restrict attention to subgame-perfect equilibrium in *pure strategies*, for the same reason as above: firms choose actions, not probability distributions over feasible actions.

<sup>18</sup>Generally speaking, Problem (1.3) is a *mathematical program with equilibrium constraints* (Luo et al., 1996).

is a pair  $(\mathbf{X}^*, \mathbf{p}^*) \in \text{graph } \mathfrak{C}(\boldsymbol{\tau}(\cdot), \mathbf{c}^U(\cdot)) \subset \mathcal{X}^J \times \mathbb{R}_+^J$ , which is to say that

$$\mathbf{X}^* \in \mathcal{X}^J \quad \text{and} \quad \mathbf{p}^* \in \mathfrak{C}(\boldsymbol{\tau}(\mathbf{X}^*), \mathbf{c}^U(\mathbf{X}^*)),$$

that solves Problem (1.3) for each  $f \in \mathbb{N}(F)$ .

Extrapolating this definition to more than two stages is conceptually straightforward. In fact, two-stage design and pricing is just a first step towards modeling firms that understand and account for *all* the future consequences of the strategic decisions they must make today. Economists now often consider “dynamic” game-theoretic models of markets, in which firms solve coupled *dynamic programs* instead of coupled optimization problems. In dynamic models where prices are a strategic variable, it is often assumed these prices are set, myopically, in Bertrand-Nash equilibrium (Ericson and Pakes, 1995; Gowrisankaran and Town, 1997; Bajari et al., 2007).<sup>19</sup>

Bertrand-Nash equilibrium prices thus play a fundamental role in Problem (1.3), as well as its extensions beyond two stages. Particularly, if equilibrium prices exist and are unique then we can view the equilibrium constraint in Problem (1.3) as simply a function that defines an intermediate variable,  $\mathbf{p}$ . This is emphasized in the theoretical economic literature studying imperfect competition using multi-stage games: almost all such studies characterize equilibrium prices as a first step in the analysis, generally making enough assumptions to guarantee that equilibrium not only exists and is unique, but has a convenient closed-form expression. So long as Problem (1.3), or an extension of it, is to play a role in modeling the impact of regulatory policies including the effects of competition, a more general understanding of Bertrand-Nash equilibrium prices will be required.

## 1.3 The Existing Understanding of Equilibrium Prices

In the previous sections, we have argued that (i) a game-theoretic theory of competitive product design and pricing under regulation must be a component of the analysis of regulatory policy in differentiated product markets and (ii) equilibrium pricing is a fundamental component of a game-theoretic models of competitive product design and pricing (whether under regulation or not). There remain two gaps in the existing understanding of equilibrium prices that stand in the way of applying Bertrand compe-

---

<sup>19</sup>Benkard (2004) gives an interesting example of an empirical dynamic Cournot competition model.

tion as a rigorous modeling real differentiated product markets. First, a theoretical analysis of price equilibrium, including the foundational assurances of existence and uniqueness, is insufficient for widespread application of Bertrand competition as a model of differentiated product markets. Second, the existing body of literature does not include a focused study of numerical methods for the computation of equilibrium prices. This section develops these two gaps in more detail.

### 1.3.1 Theoretical Understanding

While there are many existing theoretical results concerning price equilibrium, most from the multi-stage game literature, two obstacles preclude use of these results to inform the models used to study real differentiated product markets.<sup>20</sup> The first problem is that models built for theoretical study rarely consider multi-product firms, describe products through a single “quality” measure,<sup>21</sup> and rely on symmetries between the firms. Firms in real differentiated product markets, particularly automotive, ubiquitously offer more than one product and are heterogeneous in the number of products offered, the values of the many characteristics that describe these products, and also the costs with which these products can be produced.<sup>22</sup> These facts are reflected in the many empirical applications of the price equilibrium concept already discussed. While there are a few theoretical studies that consider multi-product firms (e.g., [Shaked and Sutton \(1990\)](#); [Anderson and de Palma \(1992b, 2006\)](#)), these still suffer from simplistic descriptions of the resulting firm structures and product portfolios.

Second, and more importantly, the demand models used in theoretical studies have only on rare occasions been consistent with those that are used to model demand in real differentiated product markets. Most theoretical models of differentiated product markets, like many empirical models of demand, structure consumer’s purchasing decisions by assuming that these consumers choose those products that maximize some utility function  $u$ . Many theoretical models also represent preference heterogeneity in the consumer population through the inclusion of some “demographic” variable  $\theta$  (often called a “type” in this literature), on which utility depends. However it is most often implicitly assumed that individuals’ purchases can be predicted by the firms *without error* conditional on the knowledge of this demographic. Mathematically, this “perfect conditional prediction” often implies that firms’ profit functions

---

<sup>20</sup>For examples, see the references in Section 1.2.3.

<sup>21</sup>Authors in this literature use “quality” more in the sense of “value” than in the way it would probably be interpreted by engineers, marketers, or laypeople as a measure of reliability.

<sup>22</sup>Considering the second “design” stage of these games exacerbates this problem, as most engineering models of design are quite complicated.

are non-smooth, even discontinuous, greatly complicating mathematical analyses of equilibrium.

The discrete choice RUMs of demand ([McFadden, 1981](#); [Louvierre et al., 2000](#); [Train, 2003](#)) especially popular in the many econometric models of the automotive market reject the assumption that individual consumer’s purchasing decisions can be predicted exactly, even with demographic data. Mathematically, this is a consequence of including an additive stochastic error independent of demographics. No real firm can exactly predict the purchasing behavior of consumers, even with demographic information, and thus this feature of RUMs actually provides what could be a key aspect of the real decision environment firms face. Furthermore, at least for the Logit and Mixed Logit classes of RUMs considered in this work, firms’ profit functions are continuously differentiable under very weak assumptions on the demand model specification. Because Mixed Logit models are dense in the collection of RUMs ([McFadden and Train, 2000](#)), a comprehensive study of Mixed Logit models theoretically encapsulates a study of all RUMs.

There are a few existing theoretical analyses of equilibrium prices that use empirically-relevant RUMs. [Anderson and de Palma \(1992a\)](#) consider Bertrand competition with a linear in price Logit model with symmetric single-product firms and characterize the unique equilibrium prices with a fixed-point equation. [Anderson and de Palma \(1992b\)](#) also study Bertrand competition with multi-product firms under a specific (and quite restrictive) Nested Logit Model and similarly characterize equilibrium. The assumptions made in this latter analysis essentially reduce the multi-product firm problem to one with single-product firms under Logit demand. More recently [Anderson and de Palma \(2006\)](#) study a multi-product firms under a general Nested RUM and provide another good example of how reductive these analyses can be. They admit that

“empirical application[s] would have to relax the symmetry assumptions and allow firms to produce products of different qualities, allow for heterogeneity across firms, and differing costs to introducing products.”  
([Anderson and de Palma, 2006](#), pg. 98)

Specifically, a key assumption of their analysis of price equilibrium is that firms and their products are not differentiated except by prices, a dubious choice for a theoretical model of differentiated product markets and an unacceptable choice for a model of real markets. Again, [Anderson and de Palma](#) achieve success by assuming away the essential character of the multi-product firm problem, rather than uncovering tools that can address it. Theoretical analyses that rely on such assumptions cannot support

empirical applications of the Bertrand pricing model under RUM demand.

Ideally, a general understanding of Bertrand-Nash equilibrium prices should begin with the conditions under which equilibrium prices exist. Without this understanding, empirical practitioners may specify RUMs that do not even have equilibrium prices.<sup>23</sup> Outside of equilibrium existence results applicable only to specific models, the current literature contains limited results regarding the conditions under which equilibrium will exist. Specifically, general existence conditions are only known for single-product firms. Nonetheless, these conditions have been used to ensure that both theoretical and empirical single-product firm models based on equilibrium prices are well-posed (Pakes and McGuire, 1994; Anderson et al., 1995; Gowrisankaran and Town, 1997; Anderson and de Palma, 2001).

Milgrom and Roberts (1990) and Caplin and Nalebuff (1991) provide two prominent examples of generic equilibrium existence proofs for generalized Bertrand competition with single-product firms that apply to the Logit RUM.<sup>24</sup> However, the methods employed in these examples cannot be extended to establish the existence of equilibrium prices for models with multi-product firms. While Milgrom and Roberts have applied “supermodularity” to prove the existence and uniqueness of single-product firm equilibrium prices under Logit, Sandor (2001) has shown that multi-product firm profit functions under linear in price utility Logit fail to be supermodular arbitrarily near equilibrium prices.<sup>25</sup> This result definitively excludes supermodularity as a route to establishing the existence of equilibrium prices with multi-product firms. Similarly, Caplin and Nalebuff’s proof relies on the quasi-concavity of the firms’ profit functions while Hanson and Martin (1996) have observed that multi-product firm profits are not quasi-concave. In order to deal directly with the existence of equilibrium prices for multi-product firms, new tools are needed.

### 1.3.2 Computation of Equilibrium Prices

Beyond requiring an assurance that equilibrium prices indeed exist, the growing prominence of applications of Bertrand competition to the study of real differentiated product markets requires that the computation of equilibrium prices receive greater attention. Most of the extant literature in econometrics justifiably focuses on the difficult task of model specification and estimation from market observations. Even

---

<sup>23</sup>In Chapter 3 we point out one example of where this has already occurred.

<sup>24</sup>These proofs were given only for linear in price utility functions. However, Milgrom and Roberts’s proof extends easily to any twice-differentiable and strictly decreasing utility function.

<sup>25</sup>We extend Sandor’s proof to any Logit model with a concave in price utility function.

though counterfactual experiments, one of the most important and frequent applications of such models, requires the computation of equilibrium prices, discussion of suitable methods for this task are rarely discussed in any detail. When taking the perspective of a particular firm (as is often done in engineering and marketing) demand (and cost) can be determined through controlled experimentation, thereby avoiding some of the complicating issues with estimating Bertrand competition models from market observations (e.g., see [Louvrier et al. \(2000\)](#)).

Some practitioners have identified Newton’s method, a generic algorithm for solving for zeros of nonlinear systems ([Ortega and Rheinboldt, 1970](#); [Dennis and Schnabel, 1996](#)), as the *de facto* approach to compute equilibrium prices; e.g., see ([Petrin, 2002](#); [Doraszelski and Draganska, 2006](#); [Jacobsen, 2006](#)).<sup>26</sup> In this setting, Newton’s method can be applied to compute solutions to the combined first-order optimality or “simultaneous stationarity” conditions. When successful these solutions can these be checked against the second order optimality conditions to ensure that they are indeed (local) equilibria. Recently, [Beresteanu and Li \(2008\)](#) have suggested that a common transformation of the first-order conditions can be used to solve for equilibrium prices.<sup>27</sup> This common form is a fixed-point equation for prices that dates back at least to [Berry et al. \(1995\)](#); [Goldberg \(1995\)](#), to which Newton’s method can also be applied.

Few novel approaches to the computation of either profit-optimal or equilibrium prices exist. [Hanson and Martin \(1996\)](#) derive a homotopy method for the computation of profit-optimizing prices for a restrictive (but important) subset of Mixed Logit models, but do not discuss equilibrium problems. [Choi et al. \(1990\)](#) suggest iterating on the pricing game’s best response correspondence (i.e., *tattonement*) to compute equilibrium prices.<sup>28</sup> This process itself requires a method for the computation of profit-optimal prices, ostensibly driven by an application of Newton’s method or [Hanson and Martin](#)’s homotopy method.

The efficiency, reliability, and accuracy with which any method computes equilibrium prices in realistic market models has yet to be explored in detail. Efficiency and reliability are important for the following reasons. In econometric-style analysis, *broadly* investigating the counterfactual implications of empirical models of differentiated product markets requires the analysis of equilibrium outcomes under suites of

---

<sup>26</sup>[Jacobsen \(2006\)](#) is somewhat more specific, referring to Broyden’s (or the “BFGS”) quasi-Newton method.

<sup>27</sup>They do not, however, make any reference to how the transformed first-order conditions should be attacked numerically.

<sup>28</sup>[CBO \(2003\)](#); [Michalek et al. \(2004\)](#); [Austin and Dinan \(2005\)](#) take the same approach to solving for equilibria in single-stage design and pricing games.



counterfactual assumptions, rather than simply one or a few specific cases. Similarly, including equilibrium computations in large-scale, complex multi-stage strategy games with regulation and engineering content requires the ability to compute equilibrium prices repeatedly under various values of equilibrium problem parameters (i.e., characteristics and costs). Both of these applications motivate the desire for efficient methods of computation. Furthermore, the same applications challenge the degree to which equilibrium prices can be known prior to computations, motivating computational methods that reliably compute equilibrium prices regardless of problem parameters, especially the analyst's initial guess of equilibrium prices.

Also required is an examination of the sensitivity of computed equilibrium prices to two necessary aspects of the procedure for computing equilibrium prices: choice of an initial condition and selection of a finite-sample approximation to a full Mixed Logit model. First, computed equilibrium prices are generally dependent on a set of initial prices that can be considered an initial guess of equilibrium. There has not yet been an examination of the influence of initial condition on the resulting computed equilibrium prices. Second, computing equilibrium prices under most empirical Mixed Logit models relies on computing equilibrium prices for a finite-sample approximation to the true Mixed Logit model. The sensitivity of computed equilibrium prices to the selection of a particular set of sampled demographic variables and/or random coefficient values has not been investigated.

Conclusions of counterfactual experiments may depend on the sensitivity of computed equilibrium prices to computational approximations. Table 1.1 reports Beresteanu and Li's results for one counterfactual experiment regarding gasoline prices. In this experiment, gas prices were held at 1999 levels and 2006 vehicle prices in equilibrium computed, thereby erasing the steady increase in gasoline prices that occurred during 2000-2006. As one might expect, lower gas prices appear to lower hybrid prices (as well as decrease hybrid sales). However, comparing the *level* of the price changes in Table 1.1 to our Figs. 4.7 and 4.8 in Chapter 4 should induce a degree of caution. For the 2005 vehicle market under the Berry et al. (1995) model, Fig. 4.8 demonstrates that  $10^6$  samples are required to generate 90% of computed equilibrium prices that varied by less than 100 USD when selecting the sample set. While we do not know how the sample set chosen influences the Beresteanu and Li model without explicit experiments, our results clearly suggest that the sample set size could be a key issue when studying counterfactual experiments that induce price changes on the order seen in Table 1.1. Even if the sample set does not influence the resulting conclusions, the variability in potential outcomes based on necessary

**Table 1.1** One counterfactual experiment result from [Beresteanu and Li \(2008\)](#). In this experiment gas prices are held constant at 1999 levels and new 2006 vehicle prices in equilibrium computed. Only price changes for hybrid models and their corresponding standard variants are reported.

|                   | Observed<br>Price<br>(2006 USD) | Counterfactual<br>Price<br>(2006 USD) | Change<br>(2006 USD) | (%)   |
|-------------------|---------------------------------|---------------------------------------|----------------------|-------|
| <b>Hybrid</b>     |                                 |                                       |                      |       |
| Ford Escape       | 29,140                          | 28,734                                | -406                 | -1.39 |
| Honda Civic       | 22,400                          | 21,662                                | -738                 | -3.29 |
| Honda Accord      | 31,540                          | 31,031                                | -509                 | -1.61 |
| Toyota Highlander | 34,430                          | 33,319                                | -1,111               | -3.23 |
| Toyota Prius      | 22,305                          | 21,407                                | -898                 | -4.03 |
| <b>Standard</b>   |                                 |                                       |                      |       |
| Ford Escape       | 21,745                          | 22,139                                | 394                  | 1.81  |
| Honda Civic       | 17,660                          | 17,514                                | -146                 | -0.83 |
| Honda Accord      | 21,725                          | 21,750                                | 25                   | 0.12  |
| Toyota Highlander | 26,535                          | 26,841                                | 306                  | 1.15  |
| Toyota Corolla    | 15,485                          | 15,428                                | -57                  | -0.37 |

computational approximations is an important issue to address in studies of this kind, currently ignored in the econometric community.

### 1.3.3 Empirical Validity

Despite the proliferation of Bertrand competition models of the automotive industry, there is little direct empirical evidence that Bertrand competition is an appropriate model. [Bresnahan \(1987\)](#) compares Bertrand competition to collusive pricing, finding positive evidence for Bertrand competition in a specific year (1955) and for more collusive pricing behavior in nearby years (1954 and 1956).<sup>29</sup> The historical period was examined because it presented a problematic observation for economists. Following Bresnahan, [Berndt et al. \(1990\)](#) constructed “simultaneous” and “leader/follower” (GM leading with Ford and Chrysler following) Cournot competition models of the automotive industry, finding that neither could be rejected for the period 1959-1983.

<sup>29</sup>[Sudhir \(2001\)](#) refines this balance between competition and collusion in Bertrand competition by demonstrating that contemporaneous automotive firms may collude more in some vehicle segments more than others.

Their specification was not tested against Bertrand competition. Outside of these few examples, authors do not report on direct tests of Bertrand competition assumptions.

Some authors suggest that the way to validate models of economic behavior is to validate behaviors that the models universally imply. For example, instead of “fitting” a Bertrand competition model to data, the analyst should take some property of behavior implied by the model – say constant markups – and statistically test the likelihood of this property in market data. Even without this level of rigor, models have been rejected because of the expectation that properties that hold under them are counterintuitive. For example, the Logit model we study in Chapter 3 is widely critiqued because the “independence of irrelevant of alternatives” property that many economists and marketers believe to be a spurious conclusion to make about real markets.

Generic properties of equilibrium prices under RUM demand are unknown. An understanding of what, if any, aspects of a Logit or Mixed Logit demand model generate intuitive equilibrium pricing behavior would aid model construction and application. Two avenues exist to developing this understanding. First, direct mathematical analysis can uncover structural properties that equilibrium prices must have under certain demand models. This can be a difficult task requiring skill, insight, and even luck. Second, broad numerical experimentation with the models can help develop an understanding of whether intuitive equilibrium pricing behavior exists. Some practitioners limit their reports on counterfactual equilibrium prices to specific results, as in [Beresteanu and Li \(2008\)](#). Others report only on aggregate measures ([Petrin, 2002](#); [Bento et al., 2005](#); [Jacobsen, 2006](#)), unintentionally obfuscating a general understanding the model’s realism.

## 1.4 Contributions Made in this Dissertation

This dissertation takes a step towards a rigorous understanding of game-theoretic models of realistic regulated differentiated product markets through the study of equilibrium pricing. We first introduce a general framework for Bertrand competition under an arbitrary RUM in Chapter 2. Chapters 3, 4, and 5 discuss three specific contributions to Bertrand competition with RUM demand.

- In Chapter 3 we prove the existence of unregulated equilibrium prices for the Logit class ([Train, 2003](#), Chapter 3) of discrete choice RUMs using a new fixed-point equation equivalent to the first-order necessary condition for equilibrium, minimal assumptions on the utility specification, and mathematical tools

from differential topology. We also use this equation to uncover some generic properties of equilibrium prices under the Logit model.

- In Chapter 4 we generalize this fixed-point equation to one of the most flexible and popular empirical forms for representing consumer demand, the class of Mixed Logit models (Train, 2003, Chapter 6), under a very weak hypothesis on the utility specification and mixing distribution. Applying Newton’s method to this fixed-point equation and a commonly used existing fixed-point equation is demonstrated to improve the efficiency and reliability of computations of unregulated equilibrium prices in large-scale and complex differentiated product market models relative to a naive application of Newton’s method directly to the first-order conditions. However, fixed-point iteration based on the new fixed-point equation proves to be the most reliable and efficient method for computing equilibrium prices for the example problems studied.
- Finally, in Chapter 5 we further extend this fixed-point approach to regulated equilibrium pricing problems with regulatory policy forms inspired by those considered for the U.S. automotive industry. A hybrid fixed-point iteration is developed for the computation of regulated equilibrium prices under standard-based policies with non-smooth regulatory costs.

To close this chapter, we provide a more developed introduction to these contributions in the form of an outline of the rest of this dissertation. Appendix A describes all of our basic mathematical notation, and Appendix ?? provides proofs of the mathematical results we present.

## **Analysis of Unregulated Equilibrium Prices Under Logit Models**

Chapter 3 focuses on equilibrium prices for multi-product firms under Logit models of demand. We present a new approach to studying these problems that does not require restrictive assumptions on product characteristics, firm homogeneity or symmetry, or product costs. Our approach extends in straightforward ways to most Generalized Extreme Value (GEV) (Train, 2003, Chapter 4) or Mixed Logit (Train, 2003, Chapter 6) models, although it is not yet known what conditions on the utility function, GEV function, and mixing distribution are required to generate results similar to those presented in Chapter 3.

In Section 3.1 we prove the existence of equilibrium prices by applying the Poincare-Hopf Theorem (Milnor, 1965; Simsek et al., 2007) to a new fixed-point equation derived from the Simultaneous Stationarity Condition. While similar expressions have been

posed in the literature,<sup>30</sup> this new form provides a fairly straightforward proof of the existence of equilibrium under very general assumptions on the utility function. A variety of examples of generally nonlinear in price utility functions support our definitions and assumptions. We also demonstrate that our fixed-point equation can be used to undertake a structural investigation of equilibrium prices under the Logit model under general assumptions on the utility and cost functions. The properties we identify or, more generally, our method of analysis may be of interest to practitioners using market models based on Logit.

Subsection 3.1.1 contains our conditions for the utility function under which finite equilibrium prices exist. These conditions admit any utility function that is concave in price and finite for all finite prices – functions commonly seen in theoretical economics, econometrics, and marketing. While violation of these conditions is uncommon, it is not unprecedented and can occur for models built on economic fundamentals. Particularly we use our approach to show that Bertrand competition under Logit with a utility function first derived by [Allenby and Rossi \(1991\)](#) need not have finite expected profit maximizing prices, and hence need not have finite equilibrium prices. It is important to note that this utility function was specifically derived to model an observed market behavior related to pricing.

In Section 3.2 we apply the fixed-point equation to examine properties of equilibrium prices for Bertrand competition under Logit. Particularly, our analysis allows us to identify a counter-intuitive equilibrium outcome of profit maximizing prices under Logit: the more the population values a product’s characteristics relative to the same firm’s other offerings, the lower that product’s equilibrium markup. This result is based only on the additional, reasonable assumptions that (i) the utility function is concave in price and separable in price and characteristics and (ii) unit costs increase with the value of product characteristics. A consequence of this property is that generalized Bertand competition under Logit with conventional utility specifications cannot have fixed percentage markups as an equilibrium outcome; i.e. cost plus pricing ([Nagle and Holden, 1987](#)) is not “rationalized” by Bertrand competition under conventional Logit models. Our fixed-point equation also proves a *portfolio effect*: equilibrium prices for the same products offered by two different firms at the same costs depends on the profitability of the entire portfolio of products offered by these firms. In other words, asymmetric portfolios lead to distinct equilibrium prices for equivalent products.

---

<sup>30</sup>See, for example, [Anderson and de Palma \(1992a\)](#); [Besanko et al. \(1998\)](#).

## Computation of Unregulated Equilibrium Prices under Mixed Logit Models

In Chapter 4 we extend this new approach to the class of Mixed Logit RUMs. Specifically, Section 4.1 generalizes the fixed-point equations derived in Chapter 3 for Logit models to a broad class of Mixed Logit RUMs. This class is built upon a very general specification for the systematic utility function, restricted only by a straightforward integrability condition that should hold for any reasonable Mixed Logit model. Particularly, no additional assumptions on the firms or products are made beyond those required of the general framework for Bertrand competition, and no specific restrictions are made on the mixing distribution.

Section 4.2 compares four approaches for computing equilibrium prices built from Newton’s method and fixed-point iteration. Newton’s method can be applied directly to the first-order conditions or to either of the two fixed-point equations we derive. Fixed-point iteration, in principle possible for both fixed-point expressions, only appears convergent for the new fixed-point expression. Although fixed-point iteration can only be expected to obtain a linear convergence rate it requires far fewer floating point operations (**flops**) per iteration than variants of Newton’s method and also avoids solving large linear systems.

In Section 4.3 we introduce a model of the calendar year 2005 new vehicle market with many vehicle model variants and modified versions of Mixed Logit models estimated by Boyd and Mellman (1980) and Berry et al. (1995). Section 4.4 carefully studies the properties of computations of equilibrium prices for this market using the fixed-point iteration. The most original observation is that computations may be much more sensitive to the finite-sample approximation used computations than is currently acknowledged. In Section 4.5 we employ this model of the 2005 new vehicle market to compare the practical performance of all of the numerical approaches. We conclude that, although each of the approaches based on fixed-point characterizations of equilibrium appear to be a fairly reliable means of computing equilibrium prices, the significant additional computational burden of taking Newton steps can lead to computational times far higher than those required by the fixed-point iteration.

## Computation of Regulated Equilibrium Prices

In Chapter 5 we further extend the fixed-point approach developed in the previous sections to the case where firms face several regulatory policies debated for the U.S. automotive industry. Specifically, we focus on fuel taxes, CO<sub>2</sub> taxes levied on firms

(Michalek et al., 2004), Corporate Average Fuel Economy (CAFE) standards (NHTSA, 2008b), Corporate Average Emissions (CAE) standards (CARB, 2008), and both fixed-pivot and revenue neutral feebates on fuel consumption (Greene et al., 2005). We restrict attention to single-class versions of these policies, but address how our results generalize to multi-class regulations (as in the CAFE standards currently employed in the U.S.). To characterize regulated profit-maximizing and equilibrium prices under potentially non-differentiable regulatory costs, we adopt Clarke (1975)’s generalized first-order necessary condition and Ioffe (1979)’s second-order necessary and sufficient conditions.

This fixed-point approach can be used to provide an analytical characterization of incentives engendered by different regulatory policies. Particularly, we prove the existence of a mix-shifting incentive under both the CAFE and CAE policies for most Mixed Logit models of demand, and transliterate Kleit’s simple analysis of “shadow taxes and rebates” to the general setting of regulated Bertrand competition in the automotive market under a fairly arbitrary Mixed Logit RUM.

Another contribution to regulated equilibrium problems is the proposal of a hybrid fixed-point iteration for computing equilibrium prices in regulated Bertrand competition when the regulatory costs may not be everywhere differentiable. Based on the strong performance of the fixed-point approach developed for computing unregulated equilibrium prices, we first derive modified fixed-point equations for regulated equilibrium pricing problems with both differentiable and non-differentiable regulatory costs. Through “hybridizing” the fixed-point iteration with direct solution of the non-smooth first-order conditions we propose a method applicable to policies with regulatory costs that are non-differentiable even at equilibrium prices. This hybrid approach is justified, for single-class standards, by proving fixed-point steps near profit-maximizing prices at which regulatory costs are not differentiable exhibit a cycling behavior that is easy to detect in computations.

## Chapter 2

# A General Framework for Bertrand Competition Under an Arbitrary Random Utility Model

In this chapter we present a mathematical framework for Bertrand competition used throughout this work. This generalizes the discussion in [Baye and Kovenock \(2008\)](#) by using multi-product firms and an arbitrary Random Utility Model (RUM). Conceptually, a fixed number of firms decide on prices for a fixed set of products prior to some time period in which these prices must remain fixed. During this purchasing period, a fixed number of individuals independently choose to purchase one of the products offered by these firms, or to forgo purchase of any of these products, following a given RUM. [Verboven \(1999\)](#) describes this as a two-stage stochastic game, where in the first stage the firms choose prices and in the second stage individuals choose products to maximize their own utility after sampling, or “drawing,” from the distribution of random utilities.

The goal is to derive a widely applied first-order or necessary condition for local equilibrium, the “Simultaneous Stationarity Condition,” assuming the RUM choice probabilities are continuously differentiable in prices. We first introduce the choice probabilities and expected demands arising from an arbitrary RUM. A derivation of firms’ expected profit functions then follows from the Bertrand supply assumption that the firms “committ” to produce only what individuals demand in the second stage of the game ([Baye and Kovenock, 2008](#)). Finally, we define local equilibrium and derive the Simultaneous Stationarity Condition.



## 2.1 Random Utility Models and Demand for Products

RUMs provide a means to describe selection from a *choice set*, a collection of  $J \in \mathbb{N}$  products that individuals may choose to purchase along with a no-purchase option (or “outside good”) indexed by 0. Each product  $j \in \mathbb{N}(J)$  is characterized by its price  $p_j \in [0, \infty)$  and vector of *characteristics*  $\mathbf{y}_j \in \mathcal{Y}$ , where  $\mathcal{Y} \subset \mathbb{R}^K$  for some  $K \in \mathbb{N}$ .

The random variable  $U_{i,j}(\mathbf{y}_j, p_j)$  gives the utility individual  $i$  receives by purchasing product  $j \in \mathbb{N}(J)$ . The random variable  $U_{i,0}$  gives the utility received by *not* purchasing any of the products (i.e. “purchasing the outside good”). Conditional on the values of  $\{U_{i,0}\} \cup \{U_{i,j}(\mathbf{y}_j, p_j)\}_{j \in \mathbb{N}(J)}$ , individual  $i$  chooses the option  $j \in \{0\} \cup \mathbb{N}(J)$  with the highest utility. The choice variable  $C_i(\mathbf{Y}, \mathbf{p})$  encapsulates this selection, taking values in  $\{0\} \cup \mathbb{N}(J)$  following the distribution

$$\mathbb{P}(C_i(\mathbf{Y}, \mathbf{p}) = j) = \begin{cases} \mathbb{P}\left(U_{i,j}(\mathbf{y}_j, p_j) = \max\left\{U_{i,0}, \max_{k \in \mathbb{N}(J)} U_{i,k}(\mathbf{y}_k, p_k)\right\}\right) & \text{if } j \in \mathbb{N}(J) \\ \mathbb{P}\left(U_{i,0} = \max\left\{U_{i,0}, \max_{k \in \mathbb{N}(J)} U_{i,k}(\mathbf{y}_k, p_k)\right\}\right) & \text{if } j = 0 \end{cases}$$

The distribution of these random utilities assures that “ties” occur with probability zero. We let

$$\mathbf{U}_i(\mathbf{Y}, \mathbf{p}) = (U_{i,0}, U_{i,1}(\mathbf{Y}, \mathbf{p}), \dots, U_{i,J}(\mathbf{Y}, \mathbf{p})).$$

The following assumption is usually made.

**Assumption 2.1.1.** *For any  $(\mathbf{Y}, \mathbf{p}) \in \mathcal{Y}^J \times \mathbb{R}_+^J$  and  $i, i' \in \mathbb{N}(I)$ ,  $\mathbf{U}_i(\mathbf{Y}, \mathbf{p})$  and  $\mathbf{U}_{i'}(\mathbf{Y}, \mathbf{p})$  are independent and identically distributed.*

Under this assumption, we can drop the individual index on utilities and the choice variable, referring to simply  $\mathbf{U}(\mathbf{Y}, \mathbf{p})$  and  $C(\mathbf{Y}, \mathbf{p})$ . Note also that this does not imply that  $U_{i,j}(\mathbf{y}_j, p_j)$  and  $U_{i,k}(\mathbf{y}_k, p_k)$  are independent.

In practice, models take the form

$$U_0 = \vartheta + \mathcal{E}_0 \quad \text{and} \quad U_j(\mathbf{y}_j, p_j) = u(\mathbf{y}_j, p_j) + \mathcal{E}_j \text{ for all } j \in \mathbb{N}(J)$$

for some *utility function*  $u : \mathcal{Y} \times [0, \infty) \rightarrow \mathbb{R}$ ,  $\vartheta \in [-\infty, \infty)$  and “error” vector  $\mathcal{E} = \{\mathcal{E}_j\}_{j=0}^J$ . When  $\mathcal{E}$  is given an i.i.d. extreme value distribution (NIST/SEMATECH, 2008), we have the Logit RUM (Train, 2003, Chapter 3), the focus of Chapter 3.<sup>1</sup>

<sup>1</sup>This independence assumption on  $\mathcal{E}$  is distinct from our independence assumption on the random

Letting  $\mathcal{E}$  have a Generalized Extreme Value distribution (Bierlaire et al., 2003), we have a GEV RUM (Train, 2003, Chapter 4), or taking  $\mathcal{E}$  multivariate normal gives the Probit RUM (Train, 2003, Chapter 5). Either of these forms can have a  $\mathcal{E}$  that is not an independent vector.

More generally, we can let  $\mathcal{T}$  be a space of individual characteristics or “demographics” and define

$$U_0 = \vartheta(\Theta) + \mathcal{E}_0 \quad \text{and} \quad U_j(\mathbf{y}_j, p_j) = u(\Theta, \mathbf{y}_j, p_j) + \mathcal{E}_j \text{ for all } j \in \mathbb{N}(J)$$

where  $u : \mathcal{T} \times \mathcal{Y} \times [0, \infty) \rightarrow \mathbb{R}$ ,  $\vartheta : \mathcal{T} \rightarrow [-\infty, \infty)$  where  $\Theta$  is a  $\mathcal{T}$ -valued random variable with the distribution  $\mu$ , ostensibly representing the distribution of demographic variables over the population. With the same error distribution, we obtain a “mixed” RUM. Particularly, taking  $\mathcal{E}$  i.i.d. extreme value gives the Mixed Logit RUM class (Train, 2003, Chapter 6), the object of Chapters 4 and 5.

To define firms’ profits we must also define *demands*, the total quantity of each product purchased during the purchasing period. For this, we require the following assumption.

**Assumption 2.1.2.** *Every individual  $i \in \mathbb{N}(I)$  observes the same choice set during the purchase period.*

Under this assumption, the demand  $Q_j(\mathbf{Y}, \mathbf{p})$  for each product  $j \in \mathbb{N}(J)$  can be expressed simply as  $Q_j(\mathbf{Y}, \mathbf{p}) = \sum_{i=1}^I 1_{\{C_i(\mathbf{Y}, \mathbf{p})\}}$ , where here  $\{C_i(\mathbf{Y}, \mathbf{p})\}_{i \in \mathbb{N}(I)}$  are  $I$  i.i.d. “copies” of  $C(\mathbf{Y}, \mathbf{p})$ . The primary benefit of Assumption 2.1.2 is that  $\{Q_0(\mathbf{Y}, \mathbf{p})\} \cup \{Q_j(\mathbf{Y}, \mathbf{p})\}_{j \in \mathbb{N}(J)}$  is a multinomial family of variables with parameter  $I$  and probabilities  $\{P_j(\mathbf{Y}, \mathbf{p})\}_{j=0}^J$  (Feller, 1968; Grimmett and Stirzaker, 2001). Thus expected demands for each product are given simply by  $\mathbb{E}[Q_j(\mathbf{Y}, \mathbf{p})] = IP_j(\mathbf{Y}, \mathbf{p})$ .

A more serious implication of Assumption 2.1.2 is there must be at least  $I$  units of every product available for the individuals to choose during the purchasing period. Specifically, no product can “sell out.” This is often overlooked in analyses of differentiated product markets, and appears presents an issue for Cournot competition models using RUM demand interpreted this strictly. Specifically, if any firm commits (or is forced by capacity constraints) to only produce  $I' < I$  units of some product they offer, then Assumption 2.1.2 is violated.

---

utilities in that now it is independence *across products in the choice set*, rather than *across individuals in the population*

## 2.2 Firms, Product Portfolios, Costs, and Profits

To describe the optimal pricing problems faced by each firm we must recall the following notation. Again let  $F \in \mathbb{N}$  denote the number of firms. For each  $f \in \mathbb{N}(F)$ , there exists a set  $\mathcal{J}_f \subset \mathbb{N}(J)$  of indices that corresponds to the  $J_f = |\mathcal{J}_f|$  products offered by firm  $f$ . The collection of all these sets,  $\{\mathcal{J}_f\}_{f=1}^F$ , forms a partition of  $\mathbb{N}(J)$ . Subsequently, in writing “ $f(j)$ ” for some  $j \in \mathbb{N}(J)$ , we mean the unique  $f \in \mathbb{N}(F)$  such that  $j \in \mathcal{J}_f$ . The vector  $\mathbf{p}_f \in \mathbb{R}^{J_f}$  refers to the vector of prices of the products offered by firm  $f$ . Negative subscripts continue to denote competitor’s variables as in, for instance,  $\mathbf{p}_{-f} \in \mathbb{R}^{J_{-f}}$ , where  $J_{-f} = \sum_{g \neq f} J_g$ , is the vector of prices for products offered by all of firm  $f$ ’s competitors. Firm-specific choice probability functions are denoted by  $\mathbf{P}_f(\mathbf{p})$ .

Specifying unit and fixed costs completes the definition of firms’ profits. We assume the existence of a unit variable cost function  $c_f^U : \mathcal{Y} \rightarrow \mathbb{R}_+$  and a fixed cost function  $c_f^F : \mathfrak{F}(\mathcal{Y}) \rightarrow \mathbb{R}_+$  for all  $f \in \mathbb{N}(F)$  that depend only on the collection of product characteristics chosen by the firm. Particularly, they are independent of the quantity sold.<sup>2</sup> This assumption, while standard for the price equilibrium problem, may neglect the possible dependence of unit and fixed costs on production volumes. The random variable  $\sum_{j \in \mathcal{J}_f} c_f^U(\mathbf{y}_j) Q_j(\mathbf{Y}, \mathbf{p}) + c_f^F(\mathbf{Y}_f)$  gives the total cost firm  $f$  incurs in producing  $Q_j(\mathbf{Y}, \mathbf{p})$  units of product  $j$ , for all  $j \in \mathcal{J}_f$ . We let  $\mathbf{c}_f^U(\mathbf{Y}_f)$  be the vector of these unit costs for the products offered by firm  $f$ .

Bertrand competition entails the following “commitment” assumption on the quantities produced (Baye and Kovenock, 2008).

**Assumption 2.2.1** (Bertrand Production Assumption). *Each firm produces exactly  $Q_j(\mathbf{Y}, \mathbf{p})$  units of each product  $j \in \mathcal{J}_f$  during the purchasing period.*

This implies that the firm has no production capacity constraints that limit a firm’s ability to meet any demands that arise during the purchase period or, if there are such constraints, then consumers “order” products and delivery schedules do not impact demand.

Under Assumption 2.2.1, the random variable  $\Pi_f(\mathbf{Y}, \mathbf{p}) = \mathbf{Q}_f(\mathbf{Y}, \mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f^U(\mathbf{Y}_f)) - c_f^F(\mathbf{Y}_f)$  gives firm  $f$ ’s profits for the production period as a function of prices. Following most of the theoretical and empirical literature in both marketing

---

<sup>2</sup>Strictly speaking, the fixed costs can be affine in the quantities produced, i.e.  $c_f^F(\mathbf{q}_f, \mathbf{Y}_f) = \alpha(\mathbf{Y}_f) \mathbf{q}_f^\top \bar{\mathbf{c}}_f(\mathbf{Y}_f) + \beta(\mathbf{Y}_f)$  for some  $\alpha, \beta : \mathfrak{F}(\mathcal{Y}) \rightarrow \mathbb{R}_+$  and  $\bar{\mathbf{c}}_f : \mathfrak{F}(\mathcal{Y}) \rightarrow \mathbb{R}_+^{J_f}$ . However,  $\alpha(\mathbf{Y}_f) \bar{\mathbf{c}}_f(\mathbf{Y}_f)$  is then just a vector of unit costs.

and economics, we assume that firms take expected profits,

$$\pi_f(\mathbf{Y}, \mathbf{p}) = I\hat{\pi}_f(\mathbf{Y}, \mathbf{p}) - c_f^F(\mathbf{Y}_f) \quad \text{where} \quad \hat{\pi}_f(\mathbf{Y}, \mathbf{p}) = \mathbf{P}_f(\mathbf{Y}, \mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f^U(\mathbf{Y}_f)), \quad (2.1)$$

as the metric by which they optimize their pricing decisions in this stochastic optimization problem.

We briefly demonstrate why the unit and fixed cost functions must be independent of the quantity sold. Supposing otherwise, we let  $c_f^U(\mathbf{y}_j, Q_j(\mathbf{Y}, \mathbf{p}))$  give the unit costs to firm  $f$  for offering product  $\mathbf{y}_j$  in the market of products with characteristics  $\mathbf{Y}$  and prices  $\mathbf{p}$ . Then

$$\Pi_f(\mathbf{Y}, \mathbf{p}) = \mathbf{Q}_f(\mathbf{Y}, \mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f^U(\mathbf{Y}_f, \mathbf{Q}_f(\mathbf{Y}, \mathbf{p}))) - c_f^F(\mathbf{Y}_f)$$

Notice that profits are no longer a linear function of the demands, and thus expected profits are no longer a linear function of the expected demands. This has been overlooked in some econometric studies that regress unit costs onto nontrivial functions of quantities (Berry et al., 1995; Petrin, 2002).

Eqn. (2.1) demonstrates that neither the total firm fixed costs  $c_f^F$  nor the population size  $I$  play a role in determining the prices that maximize expected profits. Therefore we only consider the “population-normalized gross expected profits”  $\hat{\pi}_f(\mathbf{p})$ , referred to below as simply “profits”. We also consider  $\mathbf{Y}$  fixed, and cease to include this characteristic matrix as an argument. Finally, we write  $\mathbf{c}_f = \mathbf{c}_f^U$  as these are the only relevant costs for the price equilibrium problem. Henceforth we write simply

$$\hat{\pi}_f(\mathbf{p}) = \mathbf{P}_f(\mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f).$$

The following adaptation of well-known necessary conditions for the local maximization of an unconstrained, continuously differentiable function (e.g., Munkres (1991)) informs our derivation of the Simultaneous Stationarity Condition.

**Proposition 2.2.1.** *Suppose  $\mathbf{P}_f(\cdot, \mathbf{p}_{-f})$  is continuously differentiable on some open  $\mathcal{A} \subset (\mathbf{0}, \infty) \subset \mathbb{R}^{J_f}$ . If  $\mathbf{p}_f \in \mathcal{A}$  is a local maximizer of  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ , then*

$$(\nabla_f \hat{\pi}_f)(\mathbf{p}) = (D_f \mathbf{P}_f)(\mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f) + \mathbf{P}_f(\mathbf{p}) = \mathbf{0}. \quad (2.2)$$

## 2.3 Local Equilibrium and the Simultaneous Stationarity Conditions

As in much of the existing literature, our analysis relies on local conditions for optimality of prices and thus must rely on the following *local* definition of equilibrium.

**Definition 2.3.1.** A price vector  $\mathbf{p} \in [0, \infty]$  is called a local equilibrium if  $\mathbf{p}_f$  is a local maximizer of  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  for all  $f \in \mathbb{N}(F)$ . A price vector  $\mathbf{p} \in [0, \infty]$  is called an equilibrium if  $\mathbf{p}_f$  is a maximizer of  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  for all  $f \in \mathbb{N}(F)$ .

Finally, the following *Simultaneous Stationarity Condition* is a generic necessary condition for local equilibrium if the RUM choice probabilities are continuously differentiable in prices.

**Definition 2.3.2.** Let  $(\tilde{\nabla}\hat{\pi})(\mathbf{p})$  denote the “combined gradient” with components  $((\tilde{\nabla}\hat{\pi})(\mathbf{p}))_j = (D_j\hat{\pi}_{f(j)})(\mathbf{p})$ . Let  $(\tilde{D}\mathbf{P})(\mathbf{p})$  be the sparse matrix corresponding to the intra-firm price derivatives of choice probabilities; that is,

$$((\tilde{D}\mathbf{P})(\mathbf{p}))_{j,k} = \begin{cases} (D_k P_j)(\mathbf{p}) & \text{if } f(j) = f(k) \\ 0 & \text{if } f(j) \neq f(k) \end{cases}.$$

**Proposition 2.3.1** (Simultaneous Stationarity Condition). Suppose  $\mathbf{P}$  is continuously differentiable on some open  $\mathcal{A} \subset (0, \infty)$ . If  $\mathbf{p} \in \mathcal{A}$  is a local equilibrium, then

$$(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = (\tilde{D}\mathbf{P})(\mathbf{p})^\top (\mathbf{p} - \mathbf{c}) + \mathbf{P}(\mathbf{p}) = \mathbf{0}. \quad (2.3)$$

Prices satisfying Eqn. (2.3) are called “simultaneously stationary.”

We emphasize that the necessity of the Simultaneous Stationarity Condition does not depend on the RUM type, but only on the continuous differentiability of the choice probabilities (with respect to price) and the cost assumption.<sup>3</sup> Thus, Eqn. (2.3) has appeared in many different studies using alternative RUM specifications. For example: Besanko et al. (1998) apply this condition for Logit models; Goldberg (1995, 1998), Besanko et al. (1998), and Villas-Boas and Zhao (2005) apply this condition for certain Generalized Extreme Value models (particularly, Nested Logit models); and Berry et al. (1995, 2004), Nevo (2000a, 2001), Sudhir (2001), Petrin (2002), and Beresteanu and Li (2008) apply this for Mixed Logit models.

---

<sup>3</sup>Furthermore, much of this development is the same for an arbitrary demand function, rather than a RUM.

In most of these studies, Eqn. (2.3) has not been investigated beyond Proposition 2.3.1. However, Eqn. (2.3) has been consistently used through the corresponding *markup equation*  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$  where

$$\boldsymbol{\eta}(\mathbf{p}) = -(\tilde{D}\mathbf{P})(\mathbf{p})^{-\top} \mathbf{P}(\mathbf{p}) \quad (2.4)$$

assuming  $(\tilde{D}\mathbf{P})(\mathbf{p})^\top$  is nonsingular.<sup>4</sup> The markup equation  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$  is a fixed-point equation satisfied by all simultaneously stationary prices that we will derive again below in the specific context of Logit and Mixed Logit models. We further derive a new fixed-point equation for simultaneously stationary prices by using a splitting of  $(\tilde{D}\mathbf{P})(\mathbf{p})^\top$  valid under Logit and Mixed Logit models.

---

<sup>4</sup>Virtually all of the papers referenced above contain this expression.

# Chapter 3

## An Analysis of Unregulated Equilibrium Prices Under Logit Models

In this chapter, we study Bertrand competition with multi-product firms, differentiated products, and Logit models of demand. Our main result is a proof of the existence of Bertrand-Nash equilibrium prices, enabled by the derivation of new conditions on the utility function and a new fixed-point equation that characterizes equilibrium prices. This fixed-point equation also allows us to demonstrate that Logit price equilibrium cannot adequately represent multi-product pricing. One counterintuitive property of equilibrium prices under most Logit models of interest to economics is identified: a firm's more valued products must have lower markups in equilibrium. Appendix B provides various examples to highlight the conditions used to prove the existence of equilibrium prices.

### 3.1 Logit Price Equilibrium

The goal of this section is to prove the following theorem regarding equilibrium prices for generalized Bertrand competition under the Logit model (defined below, in Subsection 3.1.2).

**Theorem 3.1.1.** *Suppose the utility function (cf. Assm. 3.1.1) is finite for all finite prices, is twice continuously differentiable and strictly decreases in price, eventually decreases sufficiently quickly (cf. Defn. 3.1.2), and has sub-quadratic second derivatives (cf. Defn. 3.1.3). Let  $\zeta : [0, \infty)^J \rightarrow [0, \infty)^J$  be defined as in Defn. (3.1.5).*

*If there is an outside good there is at least one equilibrium, no equilibria with infinite or non-positive markups, and a price vector  $\mathbf{p} \in (\mathbf{c}, \infty) \subset \mathbb{R}^J$  is an equilibrium if and only if  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$ .*

*If there is not an outside good, then  $\infty$  is always an equilibrium. If  $\mathbf{p} < \infty$  is an equilibrium, then in fact  $\mathbf{c} < \mathbf{p} < \infty$  and  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$ . Conversely, if there is a solution to  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  with at least one finite price for at least two firms, then  $\mathbf{c} < \mathbf{p} < \infty$  and  $\mathbf{p}$  is an equilibrium.*

Conceptually, the fixed-point equation  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  states that markups are equal to profits plus the (local) willingness to pay for product value. Loosely speaking, this equation is derived from the simultaneous stationarity conditions Eqn. (2.3) by first factoring out the gradient of the inclusive value from  $(\tilde{D}\mathbf{P})(\mathbf{p})$ . What remains is the identity and the “contractive” component of  $(\tilde{D}\mathbf{P})(\mathbf{p})$ . This equation also states that multi-product firm pricing problems under Logit are “one-parameter” problems. In other words, true multi-product pricing does not occur under Logit.

This section provides a sequence of results that culminate in Theorem 3.1.1. We separately prove that fixed-points  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  are equilibria and that such fixed-points exist. Proving existence of a finite fixed point of  $\mathbf{c} + \boldsymbol{\zeta}(\cdot)$ , even when  $\boldsymbol{\zeta}$  is not bounded, is accomplished through a straightforward application of the Poincare’-Hopf theorem. Proving that fixed-points are in fact equilibria is slightly more involved. Generally speaking, this is accomplished by showing that profits have unique maximizers, most often proving by appealing to quasi-concavity. While the multi-product firm Logit profit functions are not quasi-concave (Hanson and Martin, 1996), under utilities with sub-quadratic second derivatives first-order stationarity of profits implies local concavity. Here the Poincare-Hopf theorem implies that any firm’s profit-optimal prices are unique (for fixed competitor’s prices), effectively circumventing the difficulties with profits that are not quasi-concave. While this establishes that fixed-point are equilibria, note that it does not necessarily imply that equilibria are unique.

That  $\infty$  is always an equilibrium when there is no outside good suggests that a general treatment cannot be extended to this case without additional bounds on  $\boldsymbol{\zeta}$ .

The key results for both profit maximization and equilibrium problems are outlined with their assumptions in Tables 3.1 and 3.2.



**Table 3.1** Assumptions required for important profit maximizations results.

|               |  |   |   |                                     |
|---------------|--|---|---|-------------------------------------|
| Assm. (3.1.1) | (a-c)  | (a-c)   | $\vartheta > -\infty$<br>(a-c)            | (a-c)                               |
| Defn. (3.1.2) |  |   | $(Dw)(p) \leq -r/p$                       | $(Dw)(p) \leq -r/p$                 |
| Defn. (3.1.3) |  | $(D^2w)(p) < (Dw)(p)^2$   |   | $(D^2w)(p) < (Dw)(p)^2$             |
|               | $\mathbf{p}_f = \mathbf{c}_f + \zeta_f(\mathbf{p})$ is necessary for local profit maximization | $\mathbf{p}_f = \mathbf{c}_f + \zeta_f(\mathbf{p})$ is sufficient for local profit maximization | Local profit maximizing prices are finite | Profit-maximizing prices are unique |

**Table 3.2** Assumptions required for important equilibrium results.

|               |  |   |                                |   |
|---------------|--|---|--------------------------------|---|
| Assm. (3.1.1) | (a-c)  | (a-c)   | $\vartheta > -\infty$<br>(a-c) | (a-c)   |
| Defn. (3.1.2) |  |   | $(Dw)(p) \leq -r/p$            | $(Dw)(p) \leq -r/p$   |
| Defn. (3.1.3) |  | $(D^2w)(p) < (Dw)(p)^2$   | $(D^2w)(p) < (Dw)(p)^2$        |   |
|               | $\mathbf{p} = \mathbf{c} + \zeta(\mathbf{p})$ is necessary for finite local equilibria | $\mathbf{p} = \mathbf{c} + \zeta(\mathbf{p})$ is sufficient for finite local equilibria | Local equilibria are           | Finite solutions to $\mathbf{p} = \mathbf{c} + \zeta(\mathbf{p})$ exist |

We first define our utility specifications and the remaining terminology in Theorem 3.1.1, followed by the specific random utilities and choice probabilities of the Logit model.

### 3.1.1 Systematic Utility Specifications

The random utility any individual receives by purchasing any particular product is parameterized by its characteristic vector and price through some function  $u : \mathcal{Y} \times [0, \infty) \rightarrow [-\infty, \infty)$ . We consider specifications of the following form.

**Assumption 3.1.1.** *There are functions  $w : \mathcal{Y} \times [0, \infty) \rightarrow (-\infty, \infty)$  and  $v : \mathcal{Y} \rightarrow (-\infty, \infty)$  such that utility can be written  $u(\mathbf{y}, p) = w(\mathbf{y}, p) + v(\mathbf{y})$ . Concerning the behavior of  $w$ , we assume that, for all  $\mathbf{y} \in \mathcal{Y}$ ,  $w(\mathbf{y}, \cdot) : [0, \infty) \rightarrow (-\infty, \infty)$  is (a) strictly decreasing, and (b) continuously differentiable on  $(0, \infty)$ . We also assume that (c)  $\lim_{p \uparrow \infty} w(\mathbf{y}, p) = -\infty$ , and subsequently set  $w(\mathbf{y}, \infty) = -\infty$ .*

Writing  $u(\mathbf{y}, p) = w(\mathbf{y}, p) + v(\mathbf{y})$  is, of course, completely general, so long as utility is defined for all  $p \in [0, \infty)$ . We use this form in order to define the “value” of a product as that component of utility that does not vary with price, and to define “separable” utilities, the most common class of utility functions used in practice.

**Definition 3.1.1.** *We say  $v(\mathbf{y})$  is the value of any product with characteristic vector  $\mathbf{y}$ , and that utility is separable in price and characteristics (or simply separable) if  $w(\mathbf{y}, p) = w(p)$  for all  $\mathbf{y} \in \mathcal{Y}$ . We call  $|(Dw)(\mathbf{y}, p)|^{-1}$  the (local) willingness to pay (for product value).*

By not explicitly requiring concave-in-price utility functions, the class formed by (a-c) is slightly larger than the norm for theoretical economics and encompasses the majority of utility functions used in the empirical literature. The assumption (c) is a natural condition that ensures that the choice probabilities vanish as prices increase without bound. A number of examples are given in Appendix B.

Caplin and Nalebuff (1991) gave a prescription on which we base our specification. In our notation, they assume that  $w(\mathbf{y}, p) = \alpha(\mathbf{y})r(\varsigma - p)$  for  $p \in [0, \varsigma)$ , where  $\varsigma \in (0, \infty)$  is intended to represent income,  $r$  is some strictly increasing and concave function, and  $\alpha(\mathbf{y}) > 0$  for all  $\mathbf{y} \in \mathcal{Y}$ ;  $w$  is not defined for  $[\varsigma, \infty]$ . This prescription implies that, where defined,  $w(\mathbf{y}, p)$  is strictly decreasing and concave in price. Caplin and Nalebuff’s framework is more general in that it does not need to be defined on all of  $[0, \infty)$ .<sup>1</sup>

---

<sup>1</sup>They also allow for *heterogeneous* preferences, which we address in Chapter 4.

Concave-in-price utilities are certainly an important special case often considered in economics. However, concavity is a stronger assumption than is required to ensure the existence of finite equilibrium prices under Logit. We define the following weaker property of the utility price derivatives to achieve this.

**Definition 3.1.2.** *We say that  $w$  eventually decreases sufficiently quickly at  $\mathbf{y} \in \mathcal{Y}$  if there exists some  $r(\mathbf{y}) > 1$  and some  $\bar{p}(\mathbf{y}) \in [0, \infty)$  such that  $(Dw)(\mathbf{y}, p) \leq -r(\mathbf{y})/p = -r(\mathbf{y})D[\log p]$  for all  $p > \bar{p}(\mathbf{y})$ . We say  $w$  itself eventually decreases sufficiently quickly if  $w$  eventually decreases sufficiently quickly at all  $\mathbf{y} \in \mathcal{Y}$ .*

The most commonly used finite utility functions, particularly strictly decreasing and concave in price utility functions, satisfy  $\lim_{p \rightarrow \infty} (Dw)(\mathbf{y}, p) < 0$ , hence eventually decrease sufficiently quickly with any  $r$ .

A distinct requirement on the second derivatives of utility is synonymous with the sufficiency of stationarity under Logit.

**Definition 3.1.3.** *Suppose  $w(\mathbf{y}, \cdot)$  is twice differentiable for all  $\mathbf{y} \in \mathcal{Y}$ . We say that  $w$  has sub-quadratic second derivatives at  $(\mathbf{y}, p) \in \mathcal{Y} \times [0, \infty)$  if  $(D^2w)(\mathbf{y}, p) < (Dw)(\mathbf{y}, p)^2$ . We say that  $w$  itself has sub-quadratic second derivatives if  $w$  has sub-quadratic second derivatives at all  $(\mathbf{y}, p) \in \mathcal{Y} \times [0, \infty)$ .*

Note that if  $(D^2w)(\mathbf{y}, p) = 0$  then, under (a),  $w$  trivially has sub-quadratic second derivatives at  $(\mathbf{y}, p)$ .

With any collection of fixed product characteristic vectors  $\{\mathbf{y}_j\}_{j=1}^J$ , we set  $w_j(p) = w(\mathbf{y}_j, p)$  and  $v_j = v(\mathbf{y}_j)$  and thus generate a collection of product-specific utility functions,  $u_j(p) = w_j(p) + v_j$ , that depend on price alone. Vector functions  $\mathbf{w} : [0, \infty]^J \rightarrow [-\infty, \infty]^J$  and  $\mathbf{u} : [0, \infty]^J \rightarrow [-\infty, \infty]^J$  are constructed from these product-specific components by taking  $(\mathbf{w}(\mathbf{p}))_j = w_j(p_j)$  and  $(\mathbf{u}(\mathbf{p}))_j = u_j(p_j)$ . In particular,  $\mathbf{u}(\mathbf{p}) = \mathbf{w}(\mathbf{p}) + \mathbf{v}$ . Firm-specific product values  $\mathbf{v}_f$  and utilities  $\mathbf{u}_f(\mathbf{p}_f) = \mathbf{w}_f(\mathbf{p}_f) + \mathbf{v}_f$  are also defined in the natural way.

### 3.1.2 Logit Choice Probabilities and Profits

The Logit model (Train, 2003, Chapter 3) takes the utility any individual receives when purchasing product  $j$  to be the random variable  $U_j(\mathbf{y}_j, p_j) = u(\mathbf{y}_j, p_j) + \mathcal{E}_j$  and the utility of the outside good to be the random variable  $U_0 = \vartheta + \mathcal{E}_0$ , where  $\mathcal{E} = \{\mathcal{E}_j\}_{j=0}^J$  is a family of i.i.d. standard extreme value variables and  $\vartheta \in [-\infty, \infty)$  is a number representing the utility of the outside good.

The i.i.d. standard extreme value specification for  $\mathcal{E}$  generates the following choice probabilities (see, e.g., Train (2003)):<sup>2</sup>

$$P_j^L(\mathbf{p}) = \frac{e^{u_j(p_j)}}{e^\vartheta + \sum_{k=1}^J e^{u_k(p_k)}}. \quad (3.1)$$

We set  $e^{-\infty} = 0$  so that when  $\vartheta = -\infty$  we recover the familiar “closed” Logit formula

$$P_j^L(\mathbf{p}) = \frac{e^{u_j(p_j)}}{\sum_{k=1}^J e^{u_k(p_k)}}.$$

We use the following basic properties of the Logit choice probabilities.

**Proposition 3.1.2.** *The following hold under (a-c):*

- (i)  $0 < P_j^L(\mathbf{p}) < 1$  for all  $\mathbf{p} \in [\mathbf{0}, \infty)$  and all  $j \in \mathbb{N}(J)$ .
- (ii)  $\mathbf{P}_f^L(\mathbf{p})^\top \mathbf{1} < 1$  for all  $\mathbf{p} \in [\mathbf{0}, \infty)$  and all  $f \in \mathbb{N}(F)$ .
- (iii) If  $\vartheta > -\infty$  and  $\mathbf{q} \in [\mathbf{0}, \infty]$ ,  $\lim_{\mathbf{p} \rightarrow \mathbf{q}} P_j^L(\mathbf{p})$  exists. Moreover,  $\lim_{\mathbf{p} \rightarrow \mathbf{q}} P_j^L(\mathbf{p}) = 0$  if  $q_j = \infty$ , and  $\mathbf{P}_f^L(\mathbf{p})^\top \mathbf{1} < 1$  for all  $\mathbf{p} \in [\mathbf{0}, \infty]$  and all  $f \in \mathbb{N}(F)$ .
- (iv) If  $\vartheta = -\infty$ , then for any  $\mathbf{x} \in \mathbb{S}(J)$ , there exists some sequence  $\{\mathbf{p}^{(n)}\}_{n \in \mathbb{N}} \subset [\mathbf{0}, \infty)$  with  $\mathbf{p}^{(n)} \rightarrow \infty$  such that  $\lim_{n \rightarrow \infty} \mathbf{P}^L(\mathbf{p}^{(n)}) = \mathbf{x}$ .

Claim (iv) amounts to the surprising technical fact that the closed Logit choice probabilities cannot be single valued on  $[\mathbf{0}, \infty]$  if extended to respect limits, and suggests that the presence of an outside good purchased with positive probability, i.e.,  $\vartheta > -\infty$ , is very important to equilibrium problems under Logit. This claim is a consequence of the following basic generalization of the “invariance of uniform price shifts” property of the linear in price utility Logit model to the class of utility functions specified by (a-c).

**Lemma 3.1.3.** *Suppose  $w$  satisfies (a-c). For any  $p \in (0, \infty)$  and each  $j \in \mathbb{N}(J)$ , define  $\chi_{j,p} : [1, \infty) \rightarrow \mathbb{R}_+$  by  $\chi_{j,p}(\lambda) = w_j^{-1}(w_j(p) - \log \lambda)$ , and define  $\chi_{\mathbf{p}} : [1, \infty) \rightarrow [\mathbf{p}, \infty)$  componentwise by  $(\chi_{\mathbf{p}}(\lambda))_j = \chi_{j,p_j}(\lambda)$ .*

- (i)  $\chi_{\mathbf{p}}(\lambda)$  is strictly increasing, with  $\lim_{\lambda \rightarrow \infty} \chi_{\mathbf{p}}(\lambda) = \infty$ .
- (ii) If  $\vartheta = -\infty$ ,  $\mathbf{P}^L$  is invariant on  $\chi_{\mathbf{p}}([1, \infty))$ ; i.e.,  $\mathbf{P}^L(\chi_{\mathbf{p}}(\lambda)) \equiv \mathbf{P}^L(\mathbf{p})$ .

---

<sup>2</sup>The equivalent formula

$$P_j^L(\mathbf{p}) = \frac{e^{(u_j(p_j) - \vartheta)}}{1 + \sum_{k=1}^J e^{(u_k(p_k) - \vartheta)}}$$

corresponding to setting  $\vartheta = 0$  is often seen in the literature and can also be used, but offers no substantial advantage to our analysis in this Chapter. This form does become more useful in analyzing Mixed Logit models.

(iii) If  $\vartheta > -\infty$ ,  $\mathbf{P}^L(\boldsymbol{\chi}_{\mathbf{p}}(\lambda))$  is strictly decreasing in  $\lambda$ , and  $\mathbf{P}^L(\boldsymbol{\chi}_{\mathbf{p}}(\lambda)) \rightarrow \mathbf{0}$  as  $\lambda \rightarrow \infty$ .

We view the invariance of the choice probabilities over sequences of prices that tend to infinity as an unacceptable property for realistic market models.<sup>3</sup> Furthermore, a general existence proof is made much easier through the introduction of an outside good with positive purchase probability.

The following form for the price derivatives of the Logit choice probabilities is used extensively.

**Proposition 3.1.4.** *If  $w$  satisfies (b), then  $\mathbf{P}^L$  is continuously differentiable for all  $\mathbf{p} \in (0, \infty)$  with*

$$\begin{aligned} (D_k P_j^L)(\mathbf{p}) &= P_j^L(\mathbf{p})(\delta_{j,k} - P_k^L(\mathbf{p}))(Dw_k)(p_k) \\ &= (\delta_{j,k} - P_j^L(\mathbf{p}))P_k^L(\mathbf{p})(Dw_k)(p_k). \end{aligned} \quad (3.2)$$

In other words,

$$\begin{aligned} (D\mathbf{P}^L)(\mathbf{p}) &= \text{diag}(\mathbf{P}^L(\mathbf{p})) (\mathbf{I} - \mathbf{1}\mathbf{P}^L(\mathbf{p})^\top) (D\mathbf{w})(\mathbf{p}) \\ &= (\mathbf{I} - \mathbf{P}^L(\mathbf{p})\mathbf{1}^\top) \text{diag}(\mathbf{P}^L(\mathbf{p}))(D\mathbf{w})(\mathbf{p}) \end{aligned}$$

and

$$\begin{aligned} (D_f \mathbf{P}^L)(\mathbf{p}) &= \text{diag}(\mathbf{P}_f^L(\mathbf{p})) (\mathbf{I} - \mathbf{1}\mathbf{P}_f^L(\mathbf{p})^\top) (D\mathbf{w}_f)(\mathbf{p}_f) \\ &= (\mathbf{I} - \mathbf{P}_f^L(\mathbf{p})\mathbf{1}^\top) \text{diag}(\mathbf{P}_f^L(\mathbf{p}))(D\mathbf{w}_f)(\mathbf{p}_f) \end{aligned} \quad (3.3)$$

When  $w$  is twice differentiable, the second derivatives of the Logit choice probabilities are given by

$$\begin{aligned} (D_l D_k P_j^L)(\mathbf{p}) &= \delta_{k,l}((D^2 w_k)(p_k) + (Dw_k)(p_k)^2)P_k^L(\mathbf{p})(\delta_{j,k} - P_j^L(\mathbf{p})) \\ &\quad + (Dw_k)(p_k)P_j^L(\mathbf{p})(2P_j^L(\mathbf{p}) - \delta_{j,k} - \delta_{j,l})P_l^L(\mathbf{p})(Dw_l)(p_l). \end{aligned} \quad (3.4)$$

### 3.1.3 Bounded and Vanishing Logit Profits

An understanding of when profits are bounded over the set of all non-negative prices is a pre-requisite to a general analysis of profit-optimal prices and corresponding price

---

<sup>3</sup>Mizuno (2003) makes explicit use of this unrealistic property in proving the existence and uniqueness of equilibrium prices under Logit with single-product firms and linear in price utilities. Additionally, while we do not provide a proof of this fact in this document, it is easy to see that this property extends beyond Logit to any GEV model without an outside good.

equilibrium. One might expect that because (c) implies that the choice probabilities vanish as prices increase without bound that profits should also, but this is not true. Below we show that  $w(\mathbf{y}, p) = -\alpha \log p$  is a specification for which profits can be unbounded even though the choice probabilities vanish.<sup>4</sup> In this article, we rely on the following property of utility functions to guarantee not only the finiteness of Logit profits, but that these profits vanish as prices increase without bound.<sup>5</sup>

**Definition 3.1.4.** *We say that  $w$  is eventually log bounded at  $\mathbf{y} \in \mathcal{Y}$  if there exists some  $r(\mathbf{y}) > 1$ ,  $\kappa(\mathbf{y})$ , and some  $\bar{p}(\mathbf{y}) \in [0, \infty)$  such that  $w(\mathbf{y}, p) \leq -r(\mathbf{y}) \log p + \kappa(\mathbf{y})$  for all  $p > \bar{p}(\mathbf{y})$ . We say  $w$  itself is eventually log bounded if  $w$  is eventually log bounded at all  $\mathbf{y} \in \mathcal{Y}$ .*

We note that if  $w$  eventually decreases sufficiently quickly then the fundamental theorem of calculus implies that  $w$  is also eventually log bounded. Appendix B contains an example demonstrating that the converse need not hold.

The following proposition establishes that functions with eventually log bounded utility functions generate bounded and vanishing Logit profits.

**Proposition 3.1.5.** *Let  $\vartheta > -\infty$ ,  $\mathbf{q} \in [\mathbf{0}, \infty]$ , and suppose that there exists  $r : \mathcal{Y} \rightarrow [1, \infty)$ ,  $\bar{p} : \mathcal{Y} \rightarrow [0, \infty)$ , and  $\kappa : \mathcal{Y} \rightarrow \mathbb{R}$  such that  $w(\mathbf{y}, p) \leq -r(\mathbf{y}) \log p + \kappa(\mathbf{y})$  for all  $p > \bar{p}(\mathbf{y})$ . Then  $\lim_{\mathbf{p} \rightarrow \mathbf{q}} \hat{\pi}_f(\mathbf{p}) < \infty$ . If in fact  $r(\mathbf{y}) : \mathcal{Y} \rightarrow (1, \infty)$ , i.e. if  $w$  is eventually log bounded, then  $\lim_{\mathbf{p} \rightarrow \mathbf{q}} \hat{\pi}_f(\mathbf{p}) = 0$  if  $\mathbf{q}_f = \infty$ .*

Appendix B contains an example demonstrating that the converse to the second claim is false. That is, bounded and vanishing Logit profits need not imply that  $w$  is eventually log bounded. If eventual log boundedness is *strongly* violated in the sense of the hypothesis in the following proposition, then profits must increase without bound as prices do.

**Proposition 3.1.6.** *Let  $\vartheta > -\infty$ . Suppose that for all  $\mathbf{y} \in \mathcal{Y}$  there exists an  $r \in (0, 1)$  such that for all  $\bar{p} \in [0, \infty)$  there exists a  $p > \bar{p}$  with  $w(\mathbf{y}, p) \geq -r \log p + \kappa$ . Then  $\lim_{\mathbf{p} \rightarrow \mathbf{q}} \hat{\pi}_f(\mathbf{p}) = \infty$  for any  $\mathbf{q} \in [\mathbf{0}, \infty]$  with any  $q_j = \infty$ ,  $j \in \mathcal{J}_f$ .*

The results above prove that profit-optimal prices are not all infinite if some competitor's product or the outside good is purchased with positive probability.

---

<sup>4</sup>We draw this example from the empirical literature: [Allenby and Rossi \(1991\)](#) derive this specification from consumer-theoretic principles to represent "asymmetric brand switching under price changes."

<sup>5</sup>The constant  $\kappa(\mathbf{y})$  is convenient, but not necessary; it is easy to show that  $w$  is eventually log bounded with  $(r(\mathbf{y}), \bar{p}(\mathbf{y}), \kappa(\mathbf{y}))$  where  $\kappa(\mathbf{y}) \neq 0$  if and only if it is so with some  $(r'(\mathbf{y}), \bar{p}'(\mathbf{y}), 0)$ .

**Proposition 3.1.7.** *Suppose  $w$  satisfies (a-c) and is eventually log bounded.*

- (i) *If  $\vartheta > -\infty$  and  $\mathbf{p}_f \in [0, \infty]$  locally maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  for any  $\mathbf{p}_{-f} \in [0, \infty]$ , then  $\mathbf{p}_f \neq \infty$ .*
- (ii) *If  $\vartheta = -\infty$  and  $\mathbf{p}_f \in [0, \infty]$  locally maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  for any  $\mathbf{p}_{-f} \in [0, \infty] \setminus \{\infty\}$ , then  $\mathbf{p}_f \neq \infty$ . However,  $\hat{\pi}_f(\cdot, \infty)$  is maximized only by  $\mathbf{p}_f = \infty$ .*

This proposition states that profit-maximizing prices are not all infinite. To show that profit maximizing prices are all finite under the same hypotheses, we must take develop other aspects of our analysis.

### 3.1.4 Fixed-Point Characterizations of Price Equilibrium

In this subsection, we characterize equilibria as a fixed-point of two maps, both of which arise directly from the Simultaneous Stationarity Condition.

Our first fixed-point characterization specializes Eqn. (3.8) using Eqn. (3.3). This is a direct generalization of the fixed-point equations derived under “constant coefficient” linear in price utility (i.e.,  $w(\mathbf{y}, p) = -\alpha$  for some  $\alpha > 0$ ) for single-product firms by Anderson and de Palma (1992a) and for multi-product firms by Besanko et al. (1998).

**Proposition 3.1.8.** *Suppose  $w$  satisfies (a-c).*

- (i)  *$(\mathbf{I} - \mathbf{1P}_f^L(\mathbf{p})^\top)^{-1}$  exists whenever  $\vartheta > -\infty$  or, if  $\vartheta = -\infty$ , when  $\mathbf{p}_{-f} \neq \infty$ , but not otherwise. Moreover,  $(\mathbf{I} - \mathbf{1P}_f^L(\mathbf{p})^\top)^{-1}$  maps positive vectors to positive vectors.*
- (ii) *If  $\mathbf{p}_f \in (0, \infty)$  locally maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ , then  $\mathbf{p}_f = \mathbf{c}_f + \boldsymbol{\eta}_f(\mathbf{p})$  where*

$$\boldsymbol{\eta}_f(\mathbf{p}) = -(\mathbf{I} - \mathbf{1P}_f^L(\mathbf{p})^\top)^{-1}(D\mathbf{w}_f)(\mathbf{p}_f)^{-1}\mathbf{1}. \quad (3.5)$$

- (iii) *If  $\mathbf{p} \in (0, \infty)$  is a local equilibrium, then Eqn. (3.5) holds for all  $f \in \mathbb{N}(F)$ .*

This characterization is easily derived by noting that

$$(D_f \mathbf{P}_f)(\mathbf{p})^\top = (D_f \mathbf{w}_f)(\mathbf{p}_f) \text{diag}(\mathbf{P}_f(\mathbf{p})) (\mathbf{I} - \mathbf{1P}_f(\mathbf{p})^\top)$$

and hence Eqn. (3.8) becomes

$$(\mathbf{I} - \mathbf{1P}_f(\mathbf{p})^\top)(\mathbf{p}_f - \mathbf{c}_f) = -[(D_f \mathbf{w}_f)(\mathbf{p}_f) \text{diag}(\mathbf{P}_f(\mathbf{p}))]^{-1} \mathbf{P}_f(\mathbf{p}) = -(D_f \mathbf{w}_f)(\mathbf{p}_f)^{-1} \mathbf{1}.$$

We derive our second fixed-point function by multiplying  $(\mathbf{I} - \mathbf{1P}_f(\mathbf{p})^\top)$  through,

to obtain

$$(\mathbf{p}_f - \mathbf{c}_f) = (\mathbf{P}_f(\mathbf{p}))^\top (\mathbf{p}_f - \mathbf{c}_f) \mathbf{1} - (D_f \mathbf{w}_f)(\mathbf{p}_f)^{-1} \mathbf{1} = \hat{\pi}_f(\mathbf{p}) \mathbf{1} - (D_f \mathbf{w}_f)(\mathbf{p}_f)^{-1} \mathbf{1}.$$

We can consider  $\mathbf{1P}_f(\mathbf{p})^\top$  the “contractive” part of  $(\mathbf{I} - \mathbf{1P}_f(\mathbf{p})^\top)$  because  $\|\mathbf{1P}_f(\mathbf{p})^\top\|_\infty \leq \|\mathbf{P}_f(\mathbf{p})\|_1 < 1$ .

**Definition 3.1.5.** Define  $\zeta : (0, \infty)^J \rightarrow (-\infty, \infty)^J$  by  $\zeta(\mathbf{p}) = \tilde{\pi}(\mathbf{p}) - (D\mathbf{w})(\mathbf{p})^{-1} \mathbf{1}$  where  $\tilde{\pi}(\mathbf{p}) \in \mathbb{R}^J$  is the vector with components  $(\tilde{\pi}(\mathbf{p}))_j = \hat{\pi}_f(\mathbf{p})$ , for  $j \in \mathcal{J}_f$ .  $\zeta$  has components  $\zeta_j(\mathbf{p}) = \hat{\pi}_f(\mathbf{p}) - (Dw_j)(p_j)^{-1}$  where  $j \in \mathcal{J}_f$ , and “intra-firm” components  $\zeta_f(\mathbf{p}) = \hat{\pi}_f(\mathbf{p}) \mathbf{1} - (D\mathbf{w}_f)(\mathbf{p}_f)^{-1} \mathbf{1}$ .

The intra-firm profit gradients can be written in terms of the maps  $\zeta_f$ . Specifically,

$$(\nabla_f \hat{\pi}_f)(\mathbf{p}) = (D\mathbf{w}_f)(\mathbf{p}_f) \text{diag}(\mathbf{P}_f^L(\mathbf{p})) (\mathbf{p}_f - \mathbf{c}_f - \zeta_f(\mathbf{p})) \quad (3.6)$$

and

$$(\tilde{\nabla} \hat{\pi})(\mathbf{p}) = (D\mathbf{w})(\mathbf{p}) \text{diag}(\mathbf{P}^L(\mathbf{p})) (\mathbf{p} - \mathbf{c} - \zeta(\mathbf{p})). \quad (3.7)$$

In other words, the fixed-point “deviations”  $\mathbf{p} - \mathbf{c} - \zeta(\mathbf{p})$  arise directly from the combined gradient by factoring out the gradient of the inclusive value.

The derivation above is encapsulated in the following result.

**Proposition 3.1.9.** Suppose  $w$  satisfies (a) and (b).

- (i) If  $\mathbf{p}_f \in (\mathbf{0}, \infty)$  locally maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ , then  $\mathbf{p}_f = \mathbf{c}_f + \zeta_f(\mathbf{p}_f, \mathbf{p}_{-f})$ .
- (ii) If  $\mathbf{p} \in (\mathbf{0}, \infty)$  is a local equilibrium, then  $\mathbf{p} = \mathbf{c} + \zeta(\mathbf{p})$ .

In Appendix B we also use Eqn. (3.6) to show that profits under Logit with  $w(\mathbf{y}, p) = -\alpha \log p$  where  $\alpha \leq 1$  (Allenby and Rossi, 1991) has no finite profit-maximizing prices.<sup>6</sup>

The positivity of markups in equilibrium is an important result known to hold for single-product firm models (Anderson and de Palma, 1992a). As might be expected, profit-maximizing and equilibrium prices are not only positive but are also greater than costs under Logit. Our proof is based on both fixed-point characterizations.

**Proposition 3.1.10.** Suppose  $w$  satisfies (a-c).

- (i) No  $\mathbf{p}_f \in [\mathbf{0}, \infty] \setminus (\mathbf{c}_f, \infty]$  maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ . Hence no  $\mathbf{p} \in [\mathbf{0}, \infty] \setminus (\mathbf{c}, \infty]$  can be a local equilibrium.

---

<sup>6</sup>Sandor (2001) has also made this observation, although it is not clear he knew this function had been derived for an empirical model.



(ii) If, in addition,  $w$  is eventually log bounded and  $\vartheta > -\infty$ , then no  $\mathbf{p}_f \in [\mathbf{0}, \infty] \setminus (\mathbf{c}_f, \infty)$  maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ . Hence no  $\mathbf{p} \in [\mathbf{0}, \infty] \setminus (\mathbf{c}, \infty)$  can be a local equilibrium.

The  $\zeta$  characterization also shows that Logit price equilibrium problems are “single-parameter problems.” Specifically,  $p_j = c_j + \zeta_j(\mathbf{p})$  can also be written as  $\psi_j(p_j) = \hat{\pi}_f(\mathbf{p})$ , where  $\psi_j(p) = p - c_j + (Dw_j)(p)^{-1}$ . We note the following characteristics of the maps  $\psi_j : \mathbb{R}_+ \rightarrow \mathbb{R}$ .

**Lemma 3.1.11.** *Suppose (a-c). (i)  $\psi_j(c_j) < 0$ . (ii) If  $w$  eventually decreases sufficiently quickly, then  $\psi_j(p) \rightarrow \infty$  as  $p \rightarrow \infty$ . (iii)  $\psi_j$  is strictly increasing if, and only if,  $w$  has sub-quadratic second derivatives.*

By this Lemma, the equation  $\psi_j(p) = \pi$  has a unique solution  $\Psi_j(\pi) > c_j$  for any  $\pi > 0$  when  $w$  eventually decreases sufficiently quickly and has sub-quadratic second derivatives. Thus if  $\pi_f > 0$  are profit-optimal profits, the stationarity condition requires  $p_j = \Psi_j(\pi_f)$  for all  $j \in \mathcal{J}_f$ . In other words, we only need profits or, equivalently, a single price for each firm to derive profit-optimal prices for *all* of a firm’s products.

We can subsequently characterize Logit equilibrium in terms of profits alone, through a fixed-point of a map from  $\mathbb{R}_+^F \rightarrow \mathbb{R}_+^F$ . Let  $\Psi : \mathbb{R}_+^F \rightarrow \mathbb{R}^J$  be defined by  $(\Psi(\boldsymbol{\pi}))_j = \Psi_j(\pi_{f(j)})$ . Let  $\hat{\boldsymbol{\pi}} : \mathbb{R}^J \rightarrow \mathbb{R}^F$  have component functions  $\hat{\pi}_f : \mathbb{R}^J \rightarrow \mathbb{R}$ . The equilibrium condition is simply  $(\hat{\boldsymbol{\pi}} \circ \Psi)(\boldsymbol{\pi}) = \hat{\boldsymbol{\pi}}(\Psi(\boldsymbol{\pi})) = \boldsymbol{\pi}$ .

### 3.1.5 Sufficiency of the Fixed-Point Characterizations

Because quasi-concavity of profits is demonstrably absent with multi-product firms under Logit (Hanson and Martin, 1996) no general approach to multi-product firm equilibrium problems can rely on quasi-concavity to establish the uniqueness of profit-maximizing prices. However, something *like* quasi-concavity is required to be able to state that solutions to the fixed-point equation  $\mathbf{p}_f = \mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p})$ , which will be shown to exist, are in fact profit maximizers. When  $w$  has sub-quadratic second derivatives Logit profits have the surprising property that the first order necessary conditions imply the second order sufficient conditions; that is, stationarity implies local concavity. This circumvents the need for quasi-concavity, as desired.

**Theorem 3.1.12.** *Suppose  $w$  satisfies (a-c) and has sub-quadratic second derivatives.*

(i) Fix  $f \in \mathbb{N}(F)$ . If  $\vartheta = -\infty$ , suppose that  $\mathbf{p}_{-f} \in [\mathbf{0}, \infty] \setminus \{\infty\}$ . Satisfaction of either fixed-point equation  $\mathbf{p}_f = \mathbf{c}_f + \boldsymbol{\eta}_f(\mathbf{p}_f, \mathbf{p}_{-f})$  or  $\mathbf{p}_f = \mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}_f, \mathbf{p}_{-f})$  is sufficient for  $\mathbf{p}_f \in (\mathbf{0}, \infty) \subset \mathbb{R}^{J_f}$  to be a local maximizer of  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ .

- (ii) Satisfaction of either fixed-point equation  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  or  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$  is sufficient for  $\mathbf{p} \in (\mathbf{0}, \infty) \subset \mathbb{R}^J$  to be a local equilibrium.
- (iii) If, in addition to the hypotheses of (i),  $w$  also eventually decreases sufficiently quickly then there is a unique finite maximizer of  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ .
- (iv) When  $w$  eventually decreases sufficiently quickly, either fixed-point equation  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  or  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$  is sufficient for  $\mathbf{p} \in (\mathbf{0}, \infty) \subset \mathbb{R}^J$  to be an equilibrium.

Claim (i) is probably the most surprising result, from which Claim (ii) follows readily. The proof of Claim (i) follows from writing out the intra-firm profit price-Hessians, which turn out to be diagonal with negative diagonal entries at any stationary prices when  $w$  has sub-quadratic second derivatives. That the profit Hessians should be diagonal when pricing optimally is a surprising aspect of Logit profits suggesting that near optimality, the intra-firm profit gradients become diagonal functions.<sup>7</sup> Again setting  $\psi_j(p) = p - c_j + (Dw_j)(p)^{-1}$ , we can write

$$(D_j \hat{\pi}_f^L)(\mathbf{p}) = (Dw_j)(p_j) P_j^L(\mathbf{p}) (\psi_j(p_j) - \hat{\pi}_f^L(\mathbf{p}))$$

to see that the intra-firm profit gradients  $(\nabla_f \hat{\pi}_f^L)(\mathbf{p})$  are a row scaling of the difference between a diagonal function,  $\boldsymbol{\psi}_f(\mathbf{p}_f)$ , and a “full” function whose Jacobian vanishes when pricing optimally, namely  $\hat{\pi}_f^L(\mathbf{p}) \mathbf{1}$ .<sup>8</sup> This implies that

$$\begin{aligned} & ((Dw_j)(p_j) P_j^L(\mathbf{p}))^{-1} (D_k D_j \hat{\pi}_f^L)(\mathbf{p}) \\ &= ((Dw_j)(p_j) P_j^L(\mathbf{p}))^{-1} D_k [(Dw_j)(p_j) P_j^L(\mathbf{p})] (\psi_j(p_j) - \hat{\pi}_f^L(\mathbf{p})) \\ & \quad + \delta_{j,k} (D_j \psi)(p_j) - (D_k \hat{\pi}_f^L)(\mathbf{p}). \end{aligned}$$

When firm  $f$  prices optimally,  $(D_k D_j \hat{\pi}_f^L)(\mathbf{p})$  and  $\psi_j(p_j) - \hat{\pi}_f^L(\mathbf{p})$  vanish, leaving

$$((Dw_j)(p_j) P_j^L(\mathbf{p}))^{-1} (D_k D_j \hat{\pi}_f^L)(\mathbf{p}) = \delta_{j,k} (D_j \psi)(p_j).$$

It is easy to see that  $\psi_j$  is strictly increasing if, and only if,  $w_j$  has sub-quadratic second derivatives.

Claim (iii) can be proved with an application of the Poincare-Hopf Theorem (Milnor, 1965, Chapter 6, pg. 35) to the negative gradient vector field of firm  $f$ 's profit function. This application requires utilities to eventually decrease sufficiently quickly in order to construct a bounded compact set with a boundary on which the

<sup>7</sup>A diagonal function is simply a function with a diagonal Jacobian.

<sup>8</sup>This function always has a singular Jacobian.

negative gradient vector field always points outward. This is a consequence of the following lemma.

**Lemma 3.1.13.** *Suppose  $w$  satisfies (a-c) and eventually decreases sufficiently quickly.*

- (i) *For any  $\mathbf{p}_{-f} \in [\mathbf{0}, \infty] \setminus \{\infty\}$ , there exists some  $\bar{\mathbf{p}}_f(\mathbf{p}_{-f}) \in (\mathbf{c}_f, \infty)$  such that  $\mathbf{p}_f > \mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}_f, \mathbf{p}_{-f})$  whenever  $\mathbf{p}_f \in [\mathbf{0}, \infty] \setminus [\mathbf{0}, \bar{\mathbf{p}}_f(\mathbf{p}_{-f})]$ ; i.e., whenever  $p_j \geq \bar{p}_j(\mathbf{p}_{-f})$ .*
- (ii) *When  $\vartheta > -\infty$ , there exists some  $\bar{\mathbf{p}} \in (\mathbf{c}, \infty)$  such that  $\mathbf{p} > \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  whenever  $\mathbf{p} \in [\mathbf{0}, \infty] \setminus [\mathbf{0}, \bar{\mathbf{p}}]$ ; i.e., whenever  $p_j \geq \bar{p}_j$ .*

*On the other hand, if for some  $j$  there exists  $\bar{p}_j$  such that  $(Dw_j)(p_j) \geq -p_j^{-1}$ , then  $p_j - c_j \leq \zeta_j(\mathbf{p})$  for all  $p_j > \bar{p}_j$ .*

By Theorem 3.1.12, Claim (i), all stationary points of the negative gradient vector field have index equal to one (Milnor, 1965) and thus, by the Poincare-Hopf Theorem, there is a unique stationary point. This proof exploits the same ideas as the abstract uniqueness condition recently given by Simsek et al. (2007) for the uniqueness of solutions in inequality constrained optimization problems under a generalized Poincare-Hopf Theorem. Theorem 3.1.12, Claim (iv), then follows from Claims (ii) and (iii) of the same theorem.

Requiring utilities to have sub-quadratic second derivatives is a condition as weak as possible in the following sense. If  $w$  strictly violates the sub-quadratic second derivatives condition at  $(\mathbf{y}_j, p_j)$  where  $\mathbf{p}_f$  makes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  stationary, then profits are minimized over changes of the  $j^{\text{th}}$  price and  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  is not maximized. Furthermore, maximization of  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  by  $\mathbf{p}_f$  requires  $w$  to have sub-quadratic second derivatives at  $(\mathbf{y}_j, p_j)$  for all  $j \in \mathcal{J}_f$ : the Hessian must be negative semi-definite, including the case where it might be zero, which implies  $w$  has sub-quadratic second derivatives at  $(\mathbf{y}_j, p_j)$  for all  $j \in \mathcal{J}_f$ . In principle, it is possible that a function  $w$  that does not have sub-quadratic second derivatives for all  $(\mathbf{y}, p) \in \mathcal{Y} \times (0, \infty)$  has unique profit-maximizing prices so long as  $w$  has sub-quadratic second derivatives at every stationary point of  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ . Seeing as most functions  $w$  of current interest in economics have sub-quadratic second derivatives everywhere, investigating this complicated degree of generality does not currently seem worthwhile.

### 3.1.6 Existence of Fixed-Points

The existence claim in Theorem 3.1.1 is a consequence of the following fixed-point existence result.

**Theorem 3.1.14.** *Let  $\vartheta > -\infty$  and  $w$  satisfy (a-c) and eventually decrease sufficiently quickly. There exists at least one  $\mathbf{p} \in (\mathbf{c}, \infty)$  such that  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$ .*

The easiest proof is obtained by again using the Poincare-Hopf Theorem, which requires only that one know that  $p_j > c_j + \zeta_j(\mathbf{p})$  for large enough  $p_j$ . This result is valid even for certain *convex* in price utilities that, despite eventually decreasing sufficiently quickly, generate unbounded  $\boldsymbol{\zeta}$ .<sup>9</sup> Applying Brouwer's Fixed Point Theorem on  $[\mathbf{0}, \infty]$  is uninformative for such utilities because  $\infty$  is a fixed-point of  $\mathbf{c} + \boldsymbol{\zeta}$  when  $\lim_{p \rightarrow \infty} (Dw)(\mathbf{y}, p) = 0$ .<sup>10</sup> Furthermore, if  $w$  is not *concave* in price, it is not clear that  $\mathbf{c} + \boldsymbol{\zeta}(\cdot)$  maps  $[\mathbf{c}, \bar{\mathbf{p}}]$  into itself, where  $\bar{\mathbf{p}}$  is given by Lemma 3.1.13.

Note also that we have not required that  $w$  have sub-quadratic second derivatives. However, this assumption is necessary to make the claim that the fixed-point whose existence is guaranteed by Theorem 3.1.14 is in fact an equilibrium.

Again, the assumption  $\vartheta > -\infty$  is essential to ensure that  $\boldsymbol{\zeta}$  can be extended as a function, rather than as a set-valued map. The natural candidate for a more general existence proof applicable to the case  $\vartheta = -\infty$  is Kakutani's Fixed Point Theorem (Kakutani, 1941; Caplin and Nalebuff, 1991). However, the following straightforward extension of Proposition 3.1.7 makes any potential application of this tool on  $[\mathbf{0}, \infty]$  uninformative.

**Proposition 3.1.15.**  *$\infty$  is always an equilibrium if  $\vartheta = -\infty$ .*

It is also impossible to extend our application of the Poincare-Hopf Theorem to this case because the minimal value of  $p_j$  for which  $p_j > c_j + \zeta_j(\mathbf{p})$  holds depends on  $\mathbf{p}_{-f}$  and grows without bound as  $\mathbf{p}_{-f} \rightarrow \infty$ . This suggests that a generic treatment of the price equilibrium existence problem for  $\vartheta = -\infty$  must rely on bounds of the values taken by  $\boldsymbol{\zeta}$ .

To close this section, we state a generalization of Sandor's claim that under Logit with linear in price utility, profits are neither supermodular nor log-supermodular *arbitrarily close to equilibrium prices* (Sandor, 2001, Chapter 4).

**Proposition 3.1.16.** *Let  $\vartheta > -\infty$  and let  $w$  satisfy (a-c), be concave in price, and be twice continuously differentiable. Suppose  $\mathbf{p}_f^* \in (\mathbf{0}, \infty)$  maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ . Then for any  $\varepsilon > 0$ , there exists a  $\mathbf{p}_f$  such that  $\|\mathbf{p}_f - \mathbf{p}_f^*\| < \varepsilon$ ,  $(D_l D_k \hat{\pi}_f)(\mathbf{p}) < 0$ , and  $(D_l D_k \log \hat{\pi}_f)(\mathbf{p}) < 0$ , where  $\mathbf{p} = (\mathbf{p}_f, \mathbf{p}_{-f})$ .*

<sup>9</sup>For example,  $-\alpha \log p$ , where  $\alpha > 1$ .  $\boldsymbol{\zeta}$  is unbounded because the price derivative of utility vanishes as  $p \rightarrow \infty$ .

<sup>10</sup>Because  $[0, \infty)$  is not compact,  $|p_j - c_j - \zeta_j(\mathbf{p})|$ ,  $p_j$ , and  $c_j + \zeta_j(\mathbf{p})$  can all be unbounded as  $p_j \rightarrow \infty$ .

This result implies that there does not exist a compact set with non-empty interior containing any equilibrium on which Logit profits are supermodular or log-supermodular, ruling out the possibility of using this general property to establish the existence of equilibrium for multi-product firms as [Milgrom and Roberts \(1990\)](#) did for single-product firms. Naturally, because supermodularity has been used to prove the existence of equilibrium prices under Logit for single-product firms, the proof relies on the fact that firms produce more than one product.

## 3.2 Structural Properties of Price Equilibrium

The goal of this section is to establish properties that the finite prices and markups of any local equilibrium must satisfy based only on properties of  $\zeta$ . We emphasize that such properties depend only on the necessity of the fixed-point equation, and not on our sufficiency and existence results. The most general result is [Corollary 3.2.1](#), which states that the difference between local profit optimal markups for two products offered by the same firm depends only on the prices and characteristics of those two products. This property is very similar to the embodiment of the “Independence of Irrelevant Alternatives” (IIA) property in Logit models.<sup>11</sup> In [Corollaries 3.2.5 through 3.2.8](#) we apply this to concave-in-price utility functions under hypotheses on the unit cost and value functions to illuminate some counterintuitive properties of equilibrium prices under Logit.

### 3.2.1 Intra-Firm Structural Properties

For this subsection, we focus on a single firm  $f \in \mathbb{N}(F)$  and derive our results as properties of locally profit-optimal prices. Naturally, these properties will be manifest in locally equilibrium prices as well. This section is also the only portion of this article in which we focus heavily on concave in price utilities, which will satisfy our existence conditions. Throughout we assume that  $w$  satisfies (a-c) and  $\mathbf{p} \in (\mathbf{c}, \infty)$  is such that  $\mathbf{p}_f$  maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ , and do not list these hypotheses explicitly in the corollaries presented in this section.

The basic observation is as follows.

---

<sup>11</sup>Particularly, that the ratio of choice probabilities depends only on the characteristics and prices of those two products ([Train, 2003](#)).

**Corollary 3.2.1.** *For any  $j, k \in \mathcal{J}_f$ ,*

$$(p_j - c_j) - (p_k - c_k) = - \left( \frac{1}{(Dw_j)(p_j)} - \frac{1}{(Dw_k)(p_k)} \right). \quad (3.8)$$

That is, the difference between locally profit optimal markups for any two products offered by a single firm depends only on the corresponding utility derivatives, and hence only on the characteristics and prices of those products.

We immediately state one obvious application, motivated by the frequent application of constant coefficient linear in price utility functions.

**Corollary 3.2.2.** *If  $w(\mathbf{y}, p) \equiv -\alpha p$  for some  $\alpha > 0$ , then profit-optimal markups are constant regardless of product costs or the value of product characteristics.*

Constant intra-firm markups have appeared as an assumption (Rossi et al., 2006; Doraszelski and Draganska, 2006), but not often proven to be an equilibrium outcome (Verboven, 1999).

The following example motivates the more general propositions on profit-optimal markups given below. Consider the quadratic in price utility  $w(\mathbf{y}, p) \equiv w(p) = -\alpha p^2$ . Then

$$(p_j - c_j) - (p_k - c_k) = \left( \frac{1}{2\alpha} \right) \left( \frac{1}{p_j} - \frac{1}{p_k} \right),$$

demonstrating that locally profit optimal markups decrease with the corresponding prices (i.e.,  $m_j > m_k$  if and only if  $p_j < p_k$ ). Rearranging and setting  $\lambda = 1/(2\alpha)$ , we obtain

$$\left( p_j - \frac{\lambda}{p_j} \right) - \left( p_k - \frac{\lambda}{p_k} \right) = c_j - c_k.$$

The function  $\eta_\lambda(p) = p - \lambda/p$  is strictly increasing in  $p$  for non-negative  $\lambda$ , and thus  $c_j > c_k$  implies  $p_j > p_k$ . Thus, *locally profit optimal prices increase with costs while the corresponding markups decrease with costs*. Additionally, we note that if  $c_j = c_k$  then  $p_j = p_k$ , even if  $\mathbf{y}_j \neq \mathbf{y}_k$ ; that is, *variations in value that do not affect willingness to pay or costs are not reflected in profit-optimal prices*.

Following this example we are motivated to re-write Eqn. (3.8) as

$$\left( p_j + \frac{1}{(Dw_j)(p_j)} \right) - \left( p_k + \frac{1}{(Dw_k)(p_k)} \right) = c_j - c_k \quad (3.9)$$

essentially because of the following important observation, applied above.

**Lemma 3.2.3.** *Set  $\varphi(\mathbf{y}, p) = p + (Dw)(\mathbf{y}, p)^{-1}$ , and suppose  $w$  is twice differentiable. Then  $w$  has sub-quadratic second derivatives if, and only if,  $\varphi(\mathbf{y}, p)$  is a strictly*

increasing function of  $p$ .

We first generalize the counterintuitive property that differences in characteristics that do not impact costs or willingness to pay do not impact prices even if they impact value.

**Corollary 3.2.4.** *Let  $w$  have sub-quadratic second derivatives. Suppose  $c_j = c_k$  and  $(Dw_j)(p) = (Dw_k)(p)$  for all  $p \in [0, \infty)$  for some  $j, k \in \mathcal{J}_f$ , even if  $\mathbf{y}_j \neq \mathbf{y}_k$ . Then  $p_j = p_k$ .*

In other words, for any separable utility with sub-quadratic second derivatives, profit-optimal prices are determined by costs, not value. One would expect that real firms would not follow this rule, charging higher markups for the more valued product.

Corollary 3.2.1 also implies the second counterintuitive property of locally profit optimal markups – that they decrease with costs – under Logit with any utility function that is both strictly concave in price and separable in price and characteristics.

**Corollary 3.2.5.** *Suppose that  $w$  is separable in price and characteristics and strictly concave in price. Then firm  $f$ 's higher marginal cost products have lower locally profit optimal markups. That is, if  $j, k \in \mathcal{J}_f$  and  $c_j > c_k$ , then  $p_j - c_j < p_k - c_k$ .*

Intuition holds that both locally profit optimal markups and costs should increase with *value*, if not costs. To make this connection explicit, we pose the only quantitative assumption on costs we make in this article:

**Assumption 3.2.1** (Value Costs Hypothesis). *More valued products cost more per unit to offer; that is,  $v(\mathbf{y}) > v(\mathbf{y}')$  implies that  $c(\mathbf{y}) > c(\mathbf{y}')$  for all  $\mathbf{y}, \mathbf{y}' \in \mathcal{Y}$ .*

Bresnahan (1987) has remarked that this is a natural condition. When considering equilibrium prices, this assumption need only be applied within firms. That is, there may be firm-specific cost functions each independently satisfying the value costs hypothesis, while the value costs hypothesis is violated across firms. This states that two distinct firms can produce a value-equivalent product at distinct unit costs without violating the results that apply this hypothesis. With this definition, we provide the following restatement of Corollary 3.2.5.

**Corollary 3.2.6.** *Suppose that  $w$  is separable in price and characteristics and strictly concave in price; suppose also that the value costs hypothesis holds. Then firm  $f$ 's higher value products have lower locally profit optimal markups. That is, if  $j, k \in \mathcal{J}_f$  and  $v_j > v_k$ , then  $p_j - c_j < p_k - c_k$ .*

Markups can increase with value when  $w$  is convex in price. Consider  $w(\mathbf{y}, p) \equiv w(p) = -\alpha \log p$ ,

$$(p_j - c_j) - (p_k - c_k) = \left(\frac{1}{\alpha}\right) (p_j - p_k).$$

and we observe that locally profit optimal markups *increase* with the corresponding prices. This implies

$$(p_j - p_k) = \left(\frac{1}{\alpha - 1}\right) (c_j - c_k).$$

Hence if  $\alpha > 1$ , locally profit optimal prices increase with costs, and locally profit optimal markups *increase* with costs. While this is a more intuitive outcome, it comes from a less intuitive utility specification.

We introduce another assumption, the “unique value hypothesis,” to further connect value with prices. As defined by [Nagle and Holden \(1987\)](#), the unique value hypothesis postulates that as a product’s combination of characteristics becomes more valued, individuals are less sensitivity to price changes. We transcribe this for our framework as follows.

**Assumption 3.2.2** (Unique Value Hypothesis). *For any  $\mathbf{y}, \mathbf{y}' \in \mathcal{Y}$ ,  $v(\mathbf{y}) > v(\mathbf{y}')$  implies*

$$|(Dw)(\mathbf{y}, p)| \leq |(Dw)(\mathbf{y}', p)| \quad \text{for all } p \in [0, \infty).$$

We first use this definition to give an example of a non-separable but convex in price utility for which markups increase with value. Consider  $w(\mathbf{y}, p) = -\alpha(\mathbf{y})p$ , where  $\alpha : \mathcal{Y} \rightarrow (0, \infty)$ . Then

$$(p_j - c(\mathbf{y}_j)) - (p_k - c(\mathbf{y}_k)) = \left(\frac{1}{\alpha(\mathbf{y}_j)} - \frac{1}{\alpha(\mathbf{y}_k)}\right),$$

and  $(p_j - c(\mathbf{y}_j)) \geq (p_k - c(\mathbf{y}_k))$  if and only if  $\alpha(\mathbf{y}_j) \leq \alpha(\mathbf{y}_k)$ . The unique value hypothesis mandates that  $v(\mathbf{y}_j) > v(\mathbf{y}_k)$  implies  $\alpha(\mathbf{y}_j) \leq \alpha(\mathbf{y}_k)$ , and hence markups do not decrease with value if this hypothesis holds. Note that this is consistent with our previous result for constant coefficient linear in price utility where  $\alpha(\mathbf{y}) \equiv \alpha \in (0, \infty)$ . Whenever  $\alpha(\mathbf{y}_j) < \alpha(\mathbf{y}_k)$ , that is whenever the unique value hypothesis holds in a non-trivial way, markups can strictly increase with value.

A related and important question is whether higher value products have higher locally profit optimal prices. By [Corollary 3.2.4](#), this cannot hold without additional hypothesis. We show that this indeed holds using the value costs and unique value hypotheses.



**Corollary 3.2.7.** *Suppose that  $w$  has sub-quadratic second derivatives and satisfies the unique value hypothesis; suppose also that the value costs hypothesis holds. Then firm  $f$ 's higher value products have higher locally profit optimal prices. That is, for any  $j, k \in \mathcal{J}_f$ ,  $v_j > v_k$  implies that  $p_j > p_k$ .*

Since any separable utility trivially satisfies our “weak” specification of the unique value hypotheses, we have the following special case.

**Corollary 3.2.8.** *Suppose that  $w$  is separable in price and characteristics and that the value costs hypothesis holds. Then firm  $f$ 's higher value products have higher locally profit optimal prices. That is,  $v_j > v_k$  implies that  $p_j > p_k$  for any  $j, k \in \mathcal{J}_f$ .*

### 3.2.2 An Inter-Firm Structural Property

Eqn. (3.8) is a special case of the following equation.

**Corollary 3.2.9.** *For any  $f, g \in \mathbb{N}(F)$ ,  $j \in \mathcal{J}_f$ , and  $k \in \mathcal{J}_g$ ,*

$$(p_j - c_j) - (p_k - c_k) = \left( \hat{\pi}_f(\mathbf{p}) - \frac{1}{(Dw_j)(p_j)} \right) - \left( \hat{\pi}_g(\mathbf{p}) - \frac{1}{(Dw_k)(p_k)} \right). \quad (3.10)$$

Mirroring our approach to studying intra-firm structural properties, it is useful to set

$$\psi_f(\mathbf{y}, p) = p + \frac{1}{(Dw)(\mathbf{y}, p)} - c_f(\mathbf{y})$$

and re-write Eqn. (3.10) as

$$\psi_f(\mathbf{y}_j, p_j) - \psi_g(\mathbf{y}_k, p_k) = \hat{\pi}_f(\mathbf{p}) - \hat{\pi}_g(\mathbf{p}). \quad (3.11)$$

This equation expresses the existence of a *portfolio effect* present in equilibrium pricing with multi-product firms and even the simplest Logit model: equilibrium prices for the same product offered at the same cost by different firms can be influenced by the profitability of other products in these firm's portfolios. Stated another way, if the other products offered by a particular firm did *not* matter in determining equilibrium prices, then we would expect  $\mathbf{y}_j = \mathbf{y}_k$  and  $c_f(\mathbf{y}_j) = c_g(\mathbf{y}_k)$  for some  $j \in \mathcal{J}_f$  and  $k \in \mathcal{J}_g$  to imply that  $p_j = p_k$ . However, Eqn. (3.11) says this is not so.

**Corollary 3.2.10.** (*Portfolio Effect.*) *Let  $w$  have sub-quadratic second derivatives and suppose that  $\mathbf{y}_j = \mathbf{y}_k$  and  $c_f(\mathbf{y}_j) = c_g(\mathbf{y}_k)$  for some  $j \in \mathcal{J}_f$  and  $k \in \mathcal{J}_g$ . Then (i)  $p_j > p_k$  if, and only if,  $\hat{\pi}_f(\mathbf{p}) > \hat{\pi}_g(\mathbf{p})$ , (ii)  $p_j < p_k$  if, and only if,  $\hat{\pi}_f(\mathbf{p}) < \hat{\pi}_g(\mathbf{p})$ , and (iii)  $p_j = p_k$  if, and only if,  $\hat{\pi}_f(\mathbf{p}) = \hat{\pi}_g(\mathbf{p})$ .*

### 3.3 Conclusions

In this chapter, we have proved the existence of equilibrium prices for Bertrand competition with multi-product firms under the Logit model without restrictive assumptions on the firms or their products. Instead of studying a particular utility function, we introduce abstract conditions on the utility function under which existence holds. Our proof circumvents fundamental obstacles to the extension of existing equilibrium proofs for single-product firms by applying the Poincare-Hopf Theorem to a new fixed-point equation derived from the Simultaneous Stationarity Condition. By invoking the conventional assumption that utility is concave in price and separable in price and characteristics along with the reasonable assumption that more valued products cost more to make per unit, we have identified a counterintuitive property of profit-maximizing prices under Logit: the more the population values a product's characteristics, the lower its profit-optimal markup. We have also used the new fixed-point characterization to prove that Logit price equilibrium problems are "single-parameter problems," even when firms offer many products.

# Chapter 4

## Computation of Unregulated Equilibrium Prices Under Mixed Logit Models of Demand

In this chapter we consider four numerical approaches for computing equilibrium prices for Bertrand competition models with fairly arbitrary Mixed Logit models of demand:

**CG-NM:** Solve  $(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \mathbf{0}$  using Newton’s method;

**$\eta$ -NM:** Solve  $\mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p}) = \mathbf{0}$  using Newton’s method (see Eqn. (3.8), or Eqn. (4.1));

**$\zeta$ -NM:** Solve  $\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p}) = \mathbf{0}$  using Newton’s method (see Eqn. (4.2));

**$\zeta$ -FPI:** Iterate  $\mathbf{p} \leftarrow \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$ .<sup>1</sup>

Solving any of the three problems  $(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \mathbf{0}$ ,  $\mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p}) = \mathbf{0}$ , and  $\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p}) = \mathbf{0}$ , though equivalent in the sense of having the same solution set, may have entirely different numerical properties. Applying Newton’s method to either fixed-point formulation instead of directly attacking the simultaneous stationarity conditions can be thought of as “nonlinearly” or “analytically” pre-conditioning the original problem  $(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \mathbf{0}$ . We study fixed-point iteration based on  $\boldsymbol{\zeta}$  because we have not observed fixed-point iteration based on  $\boldsymbol{\eta}$  to be convergent.

This chapter proceeds as follows. In Section 4.1 we derive the fixed-point expressions  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$  and  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  to virtually all Mixed Logit models. Then Section 4.2 briefly reviews the numerical methods employed; Appendix C provides more details. Section 4.3 introduces our numerical example, the computation of equilibrium vehicle prices for vehicles sold during 2005 under two Mixed Logit models from the econometric literature (Boyd and Mellman, 1980; Berry et al., 1995). Section 4.4

---

<sup>1</sup>We have not found the fixed-point iteration  $\mathbf{p} \leftarrow \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$  to be convergent.

provides a detailed analysis of the behavior of the fixed-point iteration as a method of computing equilibrium prices for 5,298 vehicles sold during 2005. Section 4.5 reports on a detailed comparison between the fixed-point iteration and a state-of-the-art variant of Newton’s method applied directly to the Simultaneous Stationarity Condition and to the fixed-point characterizations using a model with only 993 vehicles. Generally, we observe that using the fixed-point characterization can result in significant increases in reliability and efficiency. Specifically,  $\zeta$ -FPI proves to be the fastest and most reliable approach for computing equilibrium prices for the model we examine.

## 4.1 Fixed-Point Equations for Equilibrium Prices Under Mixed Logit Models

In this section we generalize our fixed-point equations for local equilibrium prices derived under Logit models to any Mixed Logit model with continuously differentiable choice probabilities. Broadly speaking, Mixed Logit models characterize choice in a population through a parameterized family of simpler Logit models (Train, 2003, Chapter 6). A random vector of “demographics”  $\Theta$ , taking values in a set  $\mathcal{T}$  with probability measure  $\mu$ , serve as the parameters. These “demographics” may be either (or both) “observed” demographics (e.g., income, age, family size etc.) or may be “unobserved” demographics (Berry et al., 1995), often also called “random coefficients” (e.g., see Nevo (2000b) or Train (2003)). We first extend our class of utility functions to one that depends on these demographic variables. Next, we derive a particular “splitting” of the matrix of choice probability derivatives that allows us to write Eqn. (2.3) as a fixed-point equation of the type derived for Logit models.

### 4.1.1 Systematic Utility Specifications

Throughout this article we use the following basic utility assumption, now a proper generalization of that introduced by Caplin and Nalebuff (1991).

**Assumption 4.1.1.** *There exists some  $\varsigma : \mathcal{T} \rightarrow (0, \infty]$ ,  $w : \mathcal{T} \times \mathcal{Y} \times [0, \infty] \rightarrow [-\infty, \infty)$ , and  $v : \mathcal{T} \times \mathcal{Y} \rightarrow (-\infty, \infty)$  satisfying  $w(\boldsymbol{\theta}, \mathbf{y}, p) > -\infty$  for all  $p < \varsigma(\boldsymbol{\theta})$  and  $w(\boldsymbol{\theta}, \mathbf{y}, p) = -\infty$  for all  $p \geq \varsigma(\boldsymbol{\theta})$  such that  $u(\boldsymbol{\theta}, \mathbf{y}, p) = w(\boldsymbol{\theta}, \mathbf{y}, p) + v(\boldsymbol{\theta}, \mathbf{y})$ . Concerning the behavior of  $w$ , we assume that for  $\mu$ -almost every (a.e.)  $\boldsymbol{\theta} \in \mathcal{T}$  and all  $\mathbf{y} \in \mathcal{Y}$ ,  $w(\boldsymbol{\theta}, \mathbf{y}, \cdot) : (0, \varsigma(\boldsymbol{\theta})) \rightarrow (-\infty, \infty)$  is (a) continuously differentiable and (b) strictly decreasing, with (c)  $w(\boldsymbol{\theta}, \mathbf{y}, p) \downarrow -\infty$  as  $p \uparrow \varsigma(\boldsymbol{\theta})$ . We also suppose there exists some  $\vartheta : \mathcal{T} \rightarrow [-\infty, \infty)$ .*

We are not aware of any utility specifications used empirically that do not fall under this assumption. Individuals in the population are assigned the random utilities  $U_j(\mathbf{y}_j, p_j) = u(\Theta, \mathbf{y}_j, p_j) + \mathcal{E}_j$  for each product  $j \in \mathbb{N}(J)$  and  $U_0 = \vartheta(\Theta) + \mathcal{E}_0$  for the outside good, where  $\mathcal{E} = \{\mathcal{E}_j\}_{j=0}^J$  is a random vector of i.i.d. standard extreme value variables.

The random variable  $\Sigma = \varsigma(\Theta)$  represents a random, “non-compensatory” pricing cutoff beyond which individuals characterized by the demographics  $\Theta$  cannot purchase a product, no matter how valued its combination of characteristics. In many models  $\Sigma$  represents income, often given a lognormal distribution to fit empirical income data. We set  $\varsigma_* = \text{ess sup}_\mu \varsigma$  and allow, but do not require,  $\varsigma_* = \infty$ . An important subclass of models – those with finitely-supported  $\mu$  – will in fact have  $\varsigma_* < \infty$ .

Note also that we have not restricted  $\mu$ , the distribution of  $\Theta$ , with Assumption 4.1.1. Important examples of  $\mu$  from the econometrics and marketing literature include empirical frequency distributions, standard continuous distributions (e.g. normal, lognormal, and  $\chi^2$ ), truncated standard continuous distributions, and finite mixtures of standard continuous distributions.

## 4.1.2 Mixed Logit Choice Probabilities and Their Derivatives

The choice probabilities conditional on the value of  $\Theta$  are those of the Logit model:

$$P_j^L(\boldsymbol{\theta}, \mathbf{p}) = \mathbb{P}(C(\mathbf{p}) = j \mid \Theta = \boldsymbol{\theta}) = \frac{e^{u_j(\boldsymbol{\theta}, p_j)}}{e^{\vartheta(\boldsymbol{\theta})} + \sum_{k=1}^J e^{u_k(\boldsymbol{\theta}, p_k)}}.$$

Since  $\mathbb{P}(C(\mathbf{p}) = j) = \int \mathbb{P}(C(\mathbf{p}) = j \mid \Theta = \boldsymbol{\theta}) d\mu(\boldsymbol{\theta})$ ,  $P_j(\mathbf{p}) = \int P_j^L(\boldsymbol{\theta}, \mathbf{p}) d\mu(\boldsymbol{\theta})$  gives the Mixed Logit choice probabilities. This “mixing” of a family of Logit choice rules parameterized by  $\boldsymbol{\theta} \in \mathcal{T}$  via the “mixing distribution”  $\mu$  provides the Mixed Logit class of RUMs its name.

The following representation of the Mixed Logit choice probability price derivatives engenders the fixed-point expression given below.

**Proposition 4.1.1.** *Let  $u$  satisfy (a-c), and suppose the Leibniz Rule (e.g., see Munkres (1991)) holds for the Mixed Logit choice probabilities; that is,  $(D_k P_j)(\mathbf{p}) = \int (D_j P_k^L)(\boldsymbol{\theta}, \mathbf{p}) d\mu(\boldsymbol{\theta})$ . Then the price derivatives of the Mixed Logit choice probabilities are given by  $(DP)(\mathbf{p}) = \Lambda(\mathbf{p}) - \Gamma(\mathbf{p})$  where  $\Lambda(\mathbf{p}) = \text{diag}(\boldsymbol{\lambda}(\mathbf{p}))$ ,  $\lambda_j(\mathbf{p}) = \int (D w_j)(\boldsymbol{\theta}, p_j) P_j^L(\boldsymbol{\theta}, \mathbf{p}) d\mu(\boldsymbol{\theta})$ , and  $\Gamma$  is the full  $J \times J$  matrix with entries  $\gamma_{j,k}(\mathbf{p}) = \int P_j^L(\boldsymbol{\theta}, \mathbf{p}) P_k^L(\boldsymbol{\theta}, \mathbf{p}) (D w_k)(\boldsymbol{\theta}, p_k) d\mu(\boldsymbol{\theta})$ . The intra-firm price derivatives of the Mixed*

Logit choice probabilities are given by  $(\tilde{D}\mathbf{P})(\mathbf{p}) = \mathbf{\Lambda}(\mathbf{p}) - \tilde{\mathbf{\Gamma}}(\mathbf{p})$  where

$$(\tilde{\mathbf{\Gamma}}(\mathbf{p}))_{j,k} = \begin{cases} \gamma_{j,k}(\mathbf{p}) & \text{if } f(j) = f(k) \\ 0 & \text{if } f(j) \neq f(k) \end{cases}$$

It is easy to see that the formula

$$(D\mathbf{P})(\mathbf{p}) = (\mathbf{I} - \mathbf{\Gamma}(\mathbf{p})\mathbf{\Lambda}(\mathbf{p})^{-1})\mathbf{\Lambda}(\mathbf{p})$$

generalizes the form given in Proposition 3.1.4 for Logit models.

Finally, we provide a simple, abstract condition on the utility specification  $(u, \vartheta)$  and demographic distribution  $\mu$  guarantees that the Leibniz Rule holds and defines continuously differentiable choice probabilities. We use this condition and the proposition that follows to validate our application of the fixed-point equation to two empirical Mixed Logit RUMs of new vehicle purchasing.

**Assumption 4.1.2.**  $|(Dw)(\cdot, \mathbf{y}, p')| e^{u(\cdot, \mathbf{y}, p') - \vartheta(\cdot)}$  is uniformly  $\mu$ -integrable for all  $p'$  in a neighborhood of any  $p \in (0, \varsigma_*)$  for all  $\mathbf{y} \in \mathcal{Y}$ ; that is, there exists some  $\varphi_{(\mathbf{y}, p)} : \mathcal{T} \rightarrow [0, \infty)$  with  $\int \varphi_{(\mathbf{y}, p)}(\boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) < \infty$ , such that  $|(Dw)(\boldsymbol{\theta}, \mathbf{y}, p')| e^{u(\boldsymbol{\theta}, \mathbf{y}, p') - \vartheta(\boldsymbol{\theta})} \leq \varphi_{(\mathbf{y}, p)}(\boldsymbol{\theta})$  for all  $p'$  in some neighborhood of  $p$ .

**Proposition 4.1.2.** *If Assumption 4.1.2 holds, then the Leibniz Rule holds for the Mixed Logit choice probabilities which are also continuously differentiable on  $(\mathbf{0}, \varsigma_*\mathbf{1})$ .*

### 4.1.3 Fixed-Point Equations

We now derive fixed-point characterizations of simultaneously stationary prices. Proposition 4.1.1 states that

$$(\tilde{D}\mathbf{P})(\mathbf{p})^\top = \mathbf{\Lambda}(\mathbf{p})(\mathbf{I} - \mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top),$$

and hence Eqn. 2.4 can be written as  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$  where

$$\boldsymbol{\eta}(\mathbf{p}) = -(\mathbf{I} - \mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top)^{-1}\mathbf{\Lambda}(\mathbf{p})^{-1}\mathbf{P}(\mathbf{p}). \quad (4.1)$$

This is an extension of the second fixed-point characterization derived under Logit, and a specialization of Eqn. (3.8).<sup>2</sup> Lemma 4.1.5 establishes that  $\mathbf{I} - \mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top$  is invertible, and thus  $\boldsymbol{\eta}$  is well-defined.

<sup>2</sup>Though we could simply write  $\boldsymbol{\eta}(\mathbf{p}) = -(\tilde{D}\mathbf{P})(\mathbf{p})^{-\top}\mathbf{P}(\mathbf{p})$ , the form given in Eqn. (4.1) has resulted in a more stable means of computing  $\boldsymbol{\eta}(\mathbf{p})$ .

We could, however, write instead

$$(\mathbf{I} - \Lambda(\mathbf{p})^{-1}\tilde{\Gamma}(\mathbf{p})^\top)(\mathbf{p} - \mathbf{c}) = (\mathbf{p} - \mathbf{c}) - \Lambda(\mathbf{p})^{-1}\tilde{\Gamma}(\mathbf{p})^\top(\mathbf{p} - \mathbf{c}) = -\Lambda(\mathbf{p})^{-1}\mathbf{P}(\mathbf{p}),$$

generating the equivalent fixed-point equation  $\mathbf{p} = \mathbf{c} + \zeta(\mathbf{p})$  where

$$\zeta(\mathbf{p}) = \Lambda(\mathbf{p})^{-1}(\tilde{\Gamma}(\mathbf{p})^\top(\mathbf{p} - \mathbf{c}) - \mathbf{P}(\mathbf{p})). \quad (4.2)$$

This generalizes the other fixed-point expression derived under Logit.

We have essentially proved the following proposition:

**Proposition 4.1.3.** *Let  $\mathbf{p} \in (\mathbf{0}, \varsigma_*\mathbf{1})$  be a vector of simultaneously stationary prices,  $u$  satisfy (a-c), and Assumption 4.1.2 hold. Then  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p}) = \mathbf{c} + \zeta(\mathbf{p})$ .*

It is important to note that these two fixed-point expressions, while algebraic reorganizations of one another, are based on *distinct* maps  $\zeta$  and  $\boldsymbol{\eta}$ .<sup>3</sup> Specifically, because  $\mathbf{c} + \boldsymbol{\eta}(\cdot)$  and  $\mathbf{c} + \zeta(\cdot)$  have identical fixed-point sets, equal to the set of zeros of  $(\tilde{\nabla}\hat{\pi})$ , both maps can be used to characterize equilibrium prices as in Chapter ???. However because these maps are distinct everywhere outside of these fixed-point sets, these two characterizations can have entirely different properties when used numerically, with either Newton's method or fixed-point iteration.

**Proposition 4.1.4.** *(i)  $\zeta$  and  $\boldsymbol{\eta}$  satisfy both*

$$\begin{aligned} \zeta(\mathbf{p}) &= (\Lambda(\mathbf{p})^{-1}\tilde{\Gamma}(\mathbf{p})^\top)(\mathbf{p} - \mathbf{c}) + (\mathbf{I} - \Lambda(\mathbf{p})^{-1}\tilde{\Gamma}(\mathbf{p})^\top)\boldsymbol{\eta}(\mathbf{p}) \\ \mathbf{p} - \mathbf{c} - \zeta(\mathbf{p}) &= (\mathbf{I} - \Lambda(\mathbf{p})^{-1}\tilde{\Gamma}(\mathbf{p})^\top)(\mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p})). \end{aligned}$$

*(ii) Together, these relations imply that  $\zeta(\mathbf{p}) = \boldsymbol{\eta}(\mathbf{p})$  if, and only if,  $(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \mathbf{0}$ .*

The following Lemma helps validate this approach.

**Lemma 4.1.5.** *Let  $\mathbf{p} \in (\mathbf{0}, \varsigma_*\mathbf{1})$  and suppose  $\vartheta : \mathcal{T} \rightarrow (-\infty, \infty)$ .*

- (i)  $\|\Lambda_f(\mathbf{p})^{-1}\Gamma_f(\mathbf{p})^\top\|_\infty < 1$  and  $\|\Lambda(\mathbf{p})^{-1}\tilde{\Gamma}(\mathbf{p})^\top\|_\infty < 1$ .*
- (ii) The matrices  $\mathbf{I} - \Lambda_f(\mathbf{p})^{-1}\Gamma_f(\mathbf{p})^\top$  and  $\mathbf{I} - \Lambda(\mathbf{p})^{-1}\tilde{\Gamma}(\mathbf{p})^\top$  are strictly diagonally dominant and nonsingular. Subsequently  $(\tilde{D}\mathbf{P})(\mathbf{p})^\top$  is strictly diagonally dominant and nonsingular.*
- (iii)  $(\mathbf{I} - \Lambda_f(\mathbf{p})^{-1}\Gamma_f(\mathbf{p})^\top)^{-1}$  maps positive vectors to positive vectors.*

---

<sup>3</sup>Of course, this holds under Logit as well as Mixed Logit, even though we did not address the question of whether the two fixed-point characterizations are distinct in Chapter ??.

(iv) If  $\max \mathbf{x} > 0$ , then  $\max (\mathbf{I} - \mathbf{\Lambda}_f(\mathbf{p})^{-1}\mathbf{\Gamma}_f(\mathbf{p})^\top)\mathbf{x} > 0$ , and if  $\min \mathbf{x} < 0$ , then  $\min (\mathbf{I} - \mathbf{\Lambda}_f(\mathbf{p})^{-1}\mathbf{\Gamma}_f(\mathbf{p})^\top)\mathbf{x} < 0$ .

Claim (iv) is used only for results in the next chapter. Claim (i), discussed further in Chapter ??, is based on the simple observation that the  $k^{\text{th}}$  row sum of  $\mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top$  equals

$$\int \left( \sum_{j \in \mathcal{J}_f(k)} P_j^l(\boldsymbol{\theta}, \mathbf{p}) \right) d\mu_{k,\mathbf{p}}(\boldsymbol{\theta}) < 1$$

where

$$d\mu_{k,\mathbf{p}}(\boldsymbol{\theta}) = \frac{|(Dw_k)(\boldsymbol{\theta})| P_k^L(\boldsymbol{\theta}, \mathbf{p}) d\mu(\boldsymbol{\theta})}{\int |(Dw_k)(\boldsymbol{\theta}')| P_k^L(\boldsymbol{\theta}', \mathbf{p}) d\mu(\boldsymbol{\theta}')} \quad (4.3)$$

While we do not provide a rigorous bound, the maximal row sum of  $\mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top$  should be reasonably bounded away from one because of the influence of the outside good and a firm's competitor's products.

Much in the same manner as under Logit, Claim (iii) allows us to prove that profit-optimal markups are positive under Mixed Logit models.

**Corollary 4.1.6.** *If  $\mathbf{p}_f \in (\mathbf{0}, \varsigma_* \mathbf{1})$  locally maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ , then  $\mathbf{p}_f > \mathbf{c}_f$ . Consequently, if  $\mathbf{p} \in (\mathbf{0}, \varsigma_* \mathbf{1})$  is a local equilibrium, then  $\mathbf{p} > \mathbf{c}$ .*

## 4.2 Numerical Methods

In this section we review and compare several approaches for computing equilibrium prices. The first approach is the fixed-point iteration  $\mathbf{p} \leftarrow \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  ( $\boldsymbol{\zeta}$ -FPI). We use this fixed-point equation both because it is easy to evaluate and appears to converge reliably. The other fixed-point expression considered in this work,  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$ , does not appear to generate a convergent fixed-point iteration. The remaining approaches make use of Newton's method, a classical technique to compute a zero of an arbitrary function  $\mathbf{F} : \mathbb{R}^J \rightarrow \mathbb{R}^J$  (Ortega and Rheinboldt, 1970; Dennis and Schnabel, 1996). To compute equilibrium prices, one may take  $\mathbf{F}(\mathbf{p}) = (\tilde{\mathbf{V}}\hat{\pi})(\mathbf{p})$  (CG-NM),  $\mathbf{F}(\mathbf{p}) = \mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p})$  ( $\boldsymbol{\eta}$ -NM), or  $\mathbf{F}(\mathbf{p}) = \mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p})$  ( $\boldsymbol{\zeta}$ -NM). In Sections 4.3 and 4.5 below we apply these methods to compute equilibrium prices for an empirical example. In this section we provide a concise review of Newton's method and introduce one state-of-the-art implementation suitable for large-scale price equilibrium problems: Viswanath's GMRES-Hookstep method (Viswanath, 2007, 2008). Appendix C contains a more detailed discussion of the fixed-point iteration and variants of Newton's method.



Before reviewing the fixed-point iteration and Newton’s method, we must identify some features common to any approach.

## 4.2.1 Preliminary Considerations

### Finite-Sample Approximations

Any method for computing equilibrium prices under Mixed Logit models faces a common obstacle: the integrals that define the choice probabilities ( $\mathbf{P}$ ) and their derivatives ( $\mathbf{\Lambda}, \tilde{\mathbf{\Gamma}}$ ) cannot be computed exactly. We employ “finite-sample” versions of the methods discussed below by drawing  $S \in \mathbb{N}$  samples from the demographic distribution and applying the method to the finite-sample model thus generated. These samples are kept fixed for *all* steps of the method and, in principle, can be generated in any way. We draw directly from the demographic distribution, although importance and quasi-random sampling (e.g., see Train (2003)) can also be employed. The Law of Large Numbers (Grimmett and Stirzaker, 2001), which states that the choice probabilities and expected profits can be made as accurate as one desires by increasing the number of samples, loosely motivates this widely-used approach to econometric analysis (e.g., see McFadden (1989) and Draganska and Jain (2004)). All numerical approaches for computing equilibrium prices described here rely on a similar convergence of simultaneously stationary prices for finite-sample approximations to simultaneously stationary prices for the full model. While we do not provide a proof of this convergence, we provide numerical evidence that it does indeed occur in our examples.

### Termination Conditions

We terminate all iterations by the condition  $\|(\tilde{\mathbf{V}}\hat{\pi})(\mathbf{p})\|_{\infty} \leq \varepsilon_T$  where  $\varepsilon_T$  is some small number (e.g.,  $10^{-6}$ ). Note that a standard application of Newton’s method to solve  $\mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p}) = \mathbf{0}$  would terminate when  $\|\mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p})\|_{\infty} \leq \varepsilon_T$ . This does not necessarily imply that  $\|(\tilde{\mathbf{V}}\hat{\pi})(\mathbf{p})\|_{\infty} \leq \varepsilon_T$ . Similarly, terminating Newton’s method for  $\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p}) = \mathbf{0}$  when  $\|\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p})\|_{\infty} \leq \varepsilon_T$  does not imply  $\|(\tilde{\mathbf{V}}\hat{\pi})(\mathbf{p})\|_{\infty} \leq \varepsilon_T$ . Because

$$(\tilde{D}\mathbf{P})(\mathbf{p})^{\top}(\mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p})) = (\tilde{\mathbf{V}}\hat{\pi})(\mathbf{p}) = \mathbf{\Lambda}(\mathbf{p})(\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p})),$$

it is easy to terminate CG-NM,  $\boldsymbol{\eta}$ -NM,  $\boldsymbol{\zeta}$ -NM, and  $\boldsymbol{\zeta}$ -FPI all when  $\|(\tilde{\mathbf{V}}\hat{\pi})(\mathbf{p})\|_{\infty} \leq \varepsilon_T$ . While this ensures a degree of consistency when comparing different methods,

$\|(\tilde{\nabla}\hat{\pi}(\mathbf{p}))\|_\infty \leq \varepsilon_T$  should always be the termination condition for price equilibrium computations.

## Second-Order Sufficiency

All of the methods studied find simultaneously stationary points, rather than local equilibria. In local equilibrium every firm’s profit Hessian,  $(D_f \nabla_f \hat{\pi}_f)(\mathbf{p})$  for firm  $f$ , is negative definite.<sup>4</sup> The formulas given in Proposition C.2.1, Appendix C give an analytical expression for  $(D_f \nabla_f \hat{\pi}_f)(\mathbf{p})$  that can be approximated directly. Cholesky factorization, rather than eigenvalue estimation, can then be used to test the negative definiteness of  $(D_f \nabla_f \hat{\pi}_f)(\mathbf{p})$  (Golub and Loan, 1996).

### 4.2.2 $\zeta$ Fixed-Point Iteration

To implement the fixed-point iteration, simply iterate the assignment  $\mathbf{p} \leftarrow \mathbf{c} + \zeta(\mathbf{p})$  where Eqn. (4.2) defines  $\zeta(\mathbf{p})$ . Integral approximations, rather than the actual computation of the step, drive the computational burden of this step. Given a price vector, utilities, and utility derivatives, computing  $\mathbf{P}(\mathbf{p})$ ,  $\mathbf{\Lambda}(\mathbf{p})$ , and  $\tilde{\mathbf{\Gamma}}(\mathbf{p})$  for a set of  $S$  samples requires  $\mathcal{O}(S \sum_{f=1}^F J_f^2)$  floating point operations (flops), while the fixed-point step itself only requires  $\mathcal{O}(\sum_{f=1}^F J_f^2)$  flops. Note that computing the fixed-point step requires an equivalent amount of work as computing the combined gradient  $(\tilde{\nabla}\hat{\pi})(\mathbf{p})$ . Furthermore, no serious obstacles to computing the fixed point step arise as  $J$  becomes large, because  $\mathbf{\Lambda}(\mathbf{p})$  is a diagonal matrix.

### 4.2.3 Some Comments on the $\eta$ Map

Applying Newton’s method to  $\mathbf{p} - \mathbf{c} - \eta(\mathbf{p})$  requires computing  $\eta(\mathbf{p})$  by solving linear systems. We have found solving  $(\mathbf{I} - \mathbf{\Lambda}(\mathbf{p})^{-1} \tilde{\mathbf{\Gamma}}(\mathbf{p})^\top) \eta(\mathbf{p}) = -\mathbf{\Lambda}(\mathbf{p})^{-1} \mathbf{P}(\mathbf{p})$  more “well behaved” than solving  $(\tilde{D}\mathbf{P})(\mathbf{p})^\top \mathbf{x} = (\mathbf{\Lambda}(\mathbf{p}) - \tilde{\mathbf{\Gamma}}(\mathbf{p})^\top) \mathbf{x} = -\mathbf{P}(\mathbf{p})$ . Note also that only the systems  $(\mathbf{I} - \mathbf{\Lambda}_f(\mathbf{p})^{-1} \mathbf{\Gamma}_f(\mathbf{p})) \mathbf{x}_f = -\mathbf{\Lambda}_f(\mathbf{p})^{-1} \mathbf{P}_f(\mathbf{p})$  for all  $f$  need be solved. If QR factorization is used to solve these systems, then computing  $\eta(\mathbf{p})$  from  $\mathbf{P}(\mathbf{p})$ ,  $\mathbf{\Lambda}(\mathbf{p})$ , and  $\tilde{\mathbf{\Gamma}}(\mathbf{p})$  requires  $\mathcal{O}(\sum_{f=1}^F J_f^3)$  flops. This is a significant increase in computational effort relative to computing  $\zeta(\mathbf{p})$  or  $(\tilde{\nabla}\hat{\pi})(\mathbf{p})$ . Additional work is required to compute  $(D\eta)(\mathbf{p})$ , if this is to be used in Newton’s method. Though it requires solving a

<sup>4</sup>This is equivalent to saying that some symmetric permutation of  $(D\tilde{\nabla}\hat{\pi})(\mathbf{p})$  has a negative definite block diagonal. Note also that this (by itself) does not imply that  $(D\tilde{\nabla}\hat{\pi})(\mathbf{p})$  has *any* definiteness whatsoever.

matrix-linear system of the type  $(\tilde{D}\mathbf{P})(\mathbf{p})(D\boldsymbol{\eta})(\mathbf{p}) = \mathbf{B}(\mathbf{p})$  (cf. Appendix C), the required matrix factorizations need only be computed once.

The diagonal dominance of  $\mathbf{I} - \boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top$ , indeed of  $(\tilde{D}\mathbf{P})(\mathbf{p})$  itself, suggests that Jacobi, Gauss-Seidel, and SOR iterations may be a relatively efficient way to compute  $\boldsymbol{\eta}$ . While we have not investigated this in detail, these classical iterations illustrate a connection between the fixed-point iteration for  $\boldsymbol{\zeta}$  and computing  $\boldsymbol{\eta}$ . Note that the Jacobi-like update

$$\mathbf{x} \leftarrow \boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top \mathbf{x} - \boldsymbol{\Lambda}(\mathbf{p})^{-1}\mathbf{P}(\mathbf{p}) = \boldsymbol{\Lambda}(\mathbf{p})^{-1}(\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top \mathbf{x} - \mathbf{P}(\mathbf{p}))$$

converges to a vector  $\mathbf{x}$  solving  $(\mathbf{I} - \boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top)\mathbf{x} = -\boldsymbol{\Lambda}(\mathbf{p})^{-1}\mathbf{P}(\mathbf{p})$ . The similarities of this update formula to the fixed-point iteration based on  $\boldsymbol{\zeta}$  are obvious. Particularly, starting this iteration at  $\mathbf{p} - \mathbf{c}$  one has a first step that is exactly the update  $\mathbf{p} \leftarrow \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$ . Continuing with further steps we deviate from  $\mathbf{p} \leftarrow \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  by compounding the update based on this  $\boldsymbol{\zeta}$ -like map *without* updating  $\mathbf{P}$ ,  $\boldsymbol{\Lambda}$ , and  $\tilde{\boldsymbol{\Gamma}}$ . If, in the spirit of a nonlinear Gauss-Seidel iteration, we were to use the  $\mathbf{p}$  obtained after the first step to update  $\mathbf{P}$ ,  $\boldsymbol{\Lambda}$ , and  $\tilde{\boldsymbol{\Gamma}}$ , and then take another step of the Jacobi-like iteration, we would have the fixed-point iteration itself.

#### 4.2.4 Newton’s Method

Solving  $\mathbf{F}(\mathbf{p}) = \mathbf{0}$  through Newton’s method involves repeatedly solving for the zeros of the first-order approximations to  $\mathbf{F}(\mathbf{p})$ . That is, one solves the *Newton system*  $(D\mathbf{F})(\mathbf{p})\mathbf{s}^N = -\mathbf{F}(\mathbf{p})$  for  $\mathbf{s}^N$ , taking the update  $\mathbf{p} \leftarrow \mathbf{p} + \mathbf{s}^N$ . It must be assumed that  $(u, \vartheta, \mu)$  forms a model sufficiently “regular” so that  $(D\mathbf{F})(\mathbf{p})$  is nonsingular and reasonably well-conditioned.

Fast local convergence (Ortega and Rheinboldt, 1970; Dennis and Schnabel, 1996) often motivates the use of Newton’s method. However this fast local convergence comes at the cost of considerably more computationally intensive steps. First, computing  $\mathbf{F}(\mathbf{p})$  on the right-hand side of the Newton system has at least the computational burden of taking a  $\boldsymbol{\zeta}$  fixed-point step. To compute the Newton step  $\mathbf{s}^N$  one must also form the Jacobian matrix  $(D\mathbf{F})(\mathbf{p})$  and solve the Newton system. Finite differences are extremely wasteful, requiring  $J$  evaluations of  $\mathbf{F}$ . Quasi-Newton methods, discussed in Appendix C, approximate the Jacobian matrix in order to reduce computational burden. In Appendix C, we derive explicit expressions for the Jacobians  $(D\tilde{\nabla}\hat{\pi})$ ,  $(D\boldsymbol{\eta})$ , and  $(D\boldsymbol{\zeta})$ .<sup>5</sup> Loosely speaking, computing  $(D\mathbf{F})(\mathbf{p})$  requires  $\mathcal{O}(SJ^2)$  flops. Solving

<sup>5</sup>Automatic differentiation is another alternative to computing these Jacobians. While this may

the Newton system with direct methods, most prominently QR factorization, requires another  $\mathcal{O}(J^3)$  flops (indirect methods are discussed explicitly below). In Section 4.3 it will be observed that each step of Newton’s method can require such an increased computational burden that fast local convergence in terms of iteration count may not lead to a faster computation of equilibrium prices.

### Global Convergence.

It has long been recognized that Newton’s method only has strong *local* convergence properties (i.e. for initial conditions near solutions), often exhibiting poor *global* convergence behavior (i.e. for “arbitrary” initial conditions). By requiring that the steps taken satisfy certain “acceptability” conditions, some global convergence behavior can be enforced. See (Dennis and Schnabel, 1996, Chapter 6).

Line search, the simplest modification to Newton’s method that facilitates global convergence (e.g., see Dennis and Schnabel (1996)), employs the update  $\mathbf{p} \leftarrow \mathbf{p} + \alpha \mathbf{s}$  for some choice of step lengths  $\alpha \in \mathbb{R}$  and any step  $\mathbf{s}$ . The choice of an “acceptable” value of  $\alpha$  is an iterative process that requires evaluating the combined gradient in each iteration, which can thus be considered equivalent to taking a fixed-point step.

Another class of methods, called “trust-region” methods, instead minimizes the 2-norm of the local linear model of  $\mathbf{F}$  subject to the constraint that steps must lie within a ball of radius  $\delta$ :

$$\min_{\|\mathbf{s}\| \leq \delta} \left( \frac{1}{2} \right) \|\mathbf{F}(\mathbf{p}) + (D\mathbf{F})(\mathbf{p})\mathbf{s}\|_2^2.$$

If the Newton step can be trusted, in the sense that it has a norm less than  $\delta$ , it solves the problem above and should be taken. The Levenberg-Marquadt method or “hookstep” and Powell’s Hybrid Method or “dogleg step” (Powell, 1970; More and Trangenstein, 1976; Dennis and Schnabel, 1996) offer two alternative courses of action when the Newton step cannot be trusted. Both have seen widespread success in the iterative solution of nonlinear equations. The hookstep method computes the true solution to this problem using the Singular Value Decomposition of  $(D\mathbf{F})(\mathbf{p})$ , and can thus add significantly to the computational burden of Newton’s method. The dogleg step involves a heuristic solution to the minimization problem, with a relatively small additional computational burden. Finding an acceptable step is again an iterative process, requiring repeated evaluations of the combined gradient. Both

---

be more efficient than finite differences, it is unclear how this could be more efficient than the direct formulae we derive and use.

are implemented by `matlab`'s `fsolve` function.

### Inexact Newton Methods.

A different theory exists for the solution of systems of nonlinear equations via Newton's method when there are “many” variables. This is driven by two reasons. First, direct methods like QR factorization may not be the most effective means to solve the Newton system when this system is large, because of computational burden and accumulation of roundoff errors. Instead, iterative solution methods are often used to solve linear systems with many variables (Golub and Loan, 1996; Trefethen and Bau, 1997). Second, when far from a solution Newton steps may often point in inaccurate directions (Pernice and Walker, 1998). Thus solving for exact Newton steps may involve wasted effort, especially when there are many variables. These ideas have led to a theory of “inexact” solutions to the Newton system, defined as any vector  $\mathbf{s}^{IN}$  that satisfies  $\|\mathbf{F}(\mathbf{p}) + (D\mathbf{F})(\mathbf{p})\mathbf{s}^{IN}\| \leq \delta\|\mathbf{F}(\mathbf{p})\|$  for some  $\delta \in (0, 1)$  (Dembo et al., 1982; Eisenstat and Walker, 1994, 1996; Pernice and Walker, 1998).

Each of the globalization strategies developed for “exact” Newton methods can be applied in the inexact context. We focus on GMRES (Saad and Schultz, 1986; Walker, 1988) a particularly strong iterative method for general linear systems that has been consistently used in the context of solving nonlinear systems.<sup>6</sup> Pernice and Walker (1998) use GMRES to solve for inexact Newton steps with a safeguarded backtracking line search (Eisenstat and Walker, 1996) to facilitate global convergence. Brown and Saad (1990) and Pawlowski et al. (2006, 2008) have derived dogleg steps suitable for GMRES Newton methods. Finally, Viswanath (2007, 2008) has derived a version of the hookstep method suitable for GMRES solution of the Newton system. Viswanath's approach requires computing the Singular Value Decomposition only of a matrix whose size is determined by the number of iterations taken by GMRES. For reasonable uses of GMRES, this is far less than the size of  $(D\mathbf{F})(\mathbf{p})$  itself and can accumulate a tremendous savings over an exact-Newton implementation of the hookstep method. Each of these approaches iterates until an acceptable step is found, and can, in principle, involve many additional evaluations of  $\mathbf{F}$  or fail to find an acceptable step.

---

<sup>6</sup>`matlab`'s `fsolve` function implements a related approach using (preconditioned) conjugate gradients applied to the normal equation for the Newton system,  $(D\mathbf{F})(\mathbf{p})^\top(D\mathbf{F})(\mathbf{p})\mathbf{s}^{IN} = -(D\mathbf{F})(\mathbf{p})^\top\mathbf{F}(\mathbf{p})$ . This requires that the Jacobian of  $\mathbf{F}$  is explicitly available. Although this holds for our problems, it may be a significant restriction for general problems. Even if the Jacobian is explicitly available, this approach increases the work by  $\mathcal{O}(NJ^2)$  flops by requiring products  $(D\mathbf{F})(\mathbf{p})^\top\mathbf{y}$  in each step of the iterative linear solver. This approach can also be less accurate: using the normal equation squares the condition number (Trefethen and Bau, 1997), and can thus risk serious degradation in solution quality.

There is an additional advantage to using iterative solvers like **GMRES** in the context of nonlinear systems: only products of the type  $(D\mathbf{F})(\mathbf{p})\mathbf{s}$  will be required to solve the Newton system for  $\mathbf{F}$  at  $\mathbf{p}$ . Such a product can be approximated by a *single* additional evaluation of  $\mathbf{F}$  in a “directional” finite difference. This is, generally speaking, much less work than it takes to approximate the entire Jacobian matrix as is required by direct methods. Only needing to program  $\mathbf{F}$  can be a significant advantage in implementation, especially for highly complex problems. Ultimately, however, whether efficiency is gained by using directional finite differences instead of computing the Jacobian matrices and using standard matrix-vector products depends on the number of steps taken by the iterative linear solver. For example, if **GMRES** takes  $N \in \mathbb{N}$  iterations to find an inexact Newton step for  $\mathbf{F}$ , computing and using the Jacobian requires  $\mathcal{O}((S + N)J^2)$  flops while directional finite differences requires  $\mathcal{O}(SN \sum_{f=1}^F J_f^2)$  flops. In Appendix C we argue that, for price equilibrium problems, these relationships will tend to make computations based on the Jacobian faster. This property is shown to hold in practice as well: in our examples, using the Jacobian can be as much as half the computation time even though **GMRES** converges in very few iterations.

### 4.3 A Numerical Example: the 2005 New Vehicle Market

In this section we present the results of equilibrium price computations in a large differentiated product market with many products and a high degree of firm and product heterogeneity: the 2005 U. S. new vehicle market. This particular market has played a central role in the development of econometric applications of the generalized Bertrand competition framework for differentiated product market analysis. We begin by describing the vehicle data used and two demand models of new vehicle purchasing taken from the econometric literature. We next discuss the results of equilibrium price computations obtained using the fixed-point iteration, providing evidence that the fixed-point iteration reliably and efficiently computes local equilibrium prices.

#### 4.3.1 Vehicle Data

We combine vehicle characteristics data from Ward’s Automotive Yearbook ([Wards, 2004-2007](#)) with average cost and transaction price data for vehicles sold during 2005 as reported by dealers to J. D. Power. These data describe vehicles offered under the

**Table 4.1** Make statistics for the 2005 dataset. Abbreviations: MK: Makes; ML: Models; MLY: Model-Year Vehicles; MLYV: Model-Year Variants.

| MK         | ML | MLY | MLYV | MK            | ML         | MLY         | MLYV        |
|------------|----|-----|------|---------------|------------|-------------|-------------|
| Acura      | 6  | 18  | 26   | Lexus         | 16         | 28          | 36          |
| Audi       | 7  | 19  | 78   | Lincoln       | 6          | 15          | 49          |
| BMW        | 30 | 66  | 84   | Mazda         | 11         | 26          | 98          |
| Buick      | 9  | 18  | 66   | Mercedes-Benz | 46         | 85          | 132         |
| Cadillac   | 10 | 27  | 62   | Mercury       | 9          | 19          | 87          |
| Chevrolet  | 27 | 81  | 760  | Mini          | 1          | 3           | 10          |
| Chrysler   | 8  | 20  | 123  | Mitsubishi    | 12         | 29          | 128         |
| Dodge      | 11 | 33  | 408  | Nissan        | 11         | 39          | 268         |
| Ford       | 22 | 61  | 933  | Oldsmobile    | 2          | 3           | 18          |
| GMC        | 16 | 41  | 443  | Pontiac       | 12         | 31          | 91          |
| Honda      | 11 | 35  | 148  | Porsche       | 3          | 9           | 43          |
| Hummer     | 1  | 1   | 1    | Saab          | 4          | 8           | 32          |
| Hyundai    | 8  | 23  | 87   | Saturn        | 9          | 15          | 31          |
| Infiniti   | 8  | 17  | 38   | Scion         | 3          | 8           | 8           |
| Isuzu      | 5  | 8   | 42   | Subaru        | 6          | 17          | 89          |
| Jaguar     | 10 | 25  | 47   | Suzuki        | 7          | 19          | 90          |
| Jeep       | 5  | 14  | 99   | Toyota        | 18         | 56          | 352         |
| Kia        | 8  | 19  | 60   | Volkswagen    | 7          | 24          | 140         |
| Land Rover | 5  | 11  | 23   | Volvo         | 9          | 22          | 68          |
|            |    |     |      | <b>Median</b> | <b>8.5</b> | <b>19.5</b> | <b>81</b>   |
|            |    |     |      | <b>Total</b>  | <b>399</b> | <b>993</b>  | <b>5298</b> |

brands or “makes” enumerated in Table 4.1. Moreover, these data describe vehicles at different resolutions. J. D. Power reports average cost and transaction price (defined as price less customer cash rebate) for 993 model-year vehicles (e.g., “2005 Ford Focus”) from dealers in every major region in the U.S. On the other hand, Ward’s provides vehicle characteristics and Manufacturer’s Suggested Retail Price (MSRP) for 5,298 model-year variants (e.g., “2005 Ford Focus ZX3 S”, “2005 Ford Focus ZX3 SE”, etc.) that correspond to some cost and price observation from the J. D. Power data, after excluding virtually all vehicles with a gross vehicle weight rating above 8,500 lbs due to an absence of fuel economy data for these vehicles. To generate an equilibrium pricing problem with 5,298 vehicles we extrapolate the J. D. Power cost data by assuming that variation in dealer costs is reflected in MSRP. Specifically, we define a model-year variant’s dealer cost as the average dealer cost for the model-year vehicle with which it is associated plus the deviation of the model-year variant specific MSRP from the average MSRP across all variants of that model-year.

### 4.3.2 Two Demand Models

To characterize demand for new vehicles, we employ modified versions of two existing models of new vehicle purchasing (Boyd and Mellman, 1980; Berry et al., 1995). Here we give a brief description of the versions of these models used in our examples below.

The utility function in the Boyd and Mellman (1980) model is linear in characteristics and price with lognormally distributed “unobserved” demographic variables (random coefficients)  $\boldsymbol{\theta}$ , has no “observed” demographic variables, and does not model an outside good (i.e.,  $\vartheta(\boldsymbol{\theta}) \equiv -\infty$ ). Specifically,

$$u(\boldsymbol{\theta}, \mathbf{y}, p) = \boldsymbol{\theta}^\top \begin{bmatrix} -p \\ \mathbf{y} \end{bmatrix}.$$

We include the vehicle characteristics length times width over height (all in inches), 60 divided by the 0-60 acceleration time (in seconds), and 100 times fuel consumption (in gallons per mile). A simple function of horsepower to weight ratio approximates 0-60 acceleration, a commonly used proxy in econometric models of the vehicle market. Due to a lack of data, we exclude several consumer reports ratings (“ride,” “handling,” and “noise”) used in the original model. Finally, we use 1980 USD as our monetary units.

The utility function in the Berry et al. (1995) model is linear in characteristics with independent and normally distributed “unobserved” demographics (random coefficients), nonlinear in income minus price, and models an outside good. Specifically, with  $\boldsymbol{\theta} = (\phi, \boldsymbol{\psi})$ ,

$$u(\phi, \boldsymbol{\psi}, \mathbf{y}, p) = \alpha \log(\phi - p) + \boldsymbol{\beta}(\boldsymbol{\psi})^\top \mathbf{y} \quad \text{and} \quad \vartheta(\phi, \boldsymbol{\psi}) = \alpha \log \phi + v_0(\boldsymbol{\psi})$$

for some price coefficient  $\alpha > 0$ , income  $\phi$ , and unobserved-demographic dependent coefficient vector  $\boldsymbol{\beta}(\boldsymbol{\psi})$ . Consistent with the original model, we take income, price, and cost to be in thousands of 1983 USD where income is lognormally distributed with a log-mean of  $10 - 3 \log 10 \approx 3.092$  and a log-standard deviation of 1 (CPS, 2007). For vehicle characteristics we include operating cost (in 10 mile increments driven per dollar spent using a fuel price of 2.50 2005 USD = 1.27 1983 USD), horsepower to weight ratio (weight is in 10 lbs.), and length times width (both in hundreds of inches). We effectively exclude a dummy variable for standard air conditioning present in the original model by assuming standard air conditioning on all vehicles. No other characteristics from the original model differ in our version.



**Table 4.2** “Worst-case” observed computation times for various sample set sizes.

|                   |       | Boyd and Mellman (1980) |        |        |        | Berry et al. (1995) |        |        |        |
|-------------------|-------|-------------------------|--------|--------|--------|---------------------|--------|--------|--------|
| Number of Samples | (#)   | $10^3$                  | $10^4$ | $10^5$ | $10^6$ | $10^3$              | $10^4$ | $10^5$ | $10^6$ |
| CPU Time          | (hrs) | 0.1                     | 2.4    | 4.4    | 65     | 0.2                 | 1.4    | 7.5    | 125    |

The following proposition ensures the validity of our treatment of equilibrium prices using the fixed-point iteration under these two demand models.

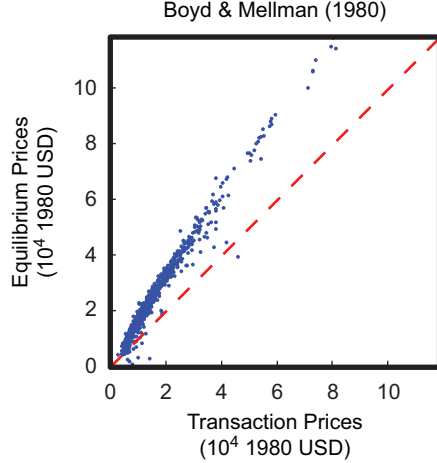
**Proposition 4.3.1.** *(i) The conclusions of Proposition 4.1.2 hold under the Boyd and Mellman (1980) model. (ii) Assumption 4.1.2, and hence the conclusions of Proposition 4.1.2, hold under the Berry et al. (1995) model so long as  $\alpha > 1$ .*

We treat the Boyd and Mellman model separately because it does not have an outside good, and hence does not match the format of Assumption 4.1.2.

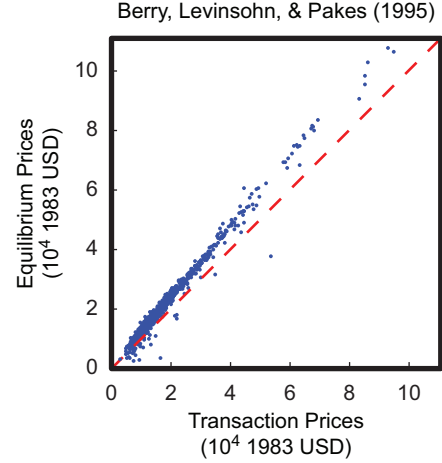
## 4.4 Computing Equilibrium Prices with the Fixed-Point Iteration

In Figs. 4.1 and 4.2 we compare equilibrium prices, computed via the fixed-point iteration, to transaction prices for a single 1,000,000 sample set. Note that in comparing the 5,298 computed equilibrium prices at the model variant level with the 993 average transaction prices at the model level, we average the computed model variant prices using the choice probabilities for all variants of the corresponding model as weights. Computed equilibrium prices under both demand models appear to over-predict average transaction prices, with the error increasing with price and smallest under the Berry et al. model. In Table 4.2 we list the longest times taken to compute equilibrium prices using the fixed-point iteration. A simple analysis of the fixed-point iteration suggests that adding samples increases the computational burden in a linear fashion, as observed in these times.

Figs. 4.3 and 4.4 demonstrate the effectiveness of the fixed-point iteration as a method of computing simultaneously stationary prices for all 5,298 vehicles under both demand models and “arbitrary” initial conditions. For the Boyd and Mellman model we test 10 initial conditions drawn uniformly from  $[\min_{j \in \mathbb{N}(J)} c_j, \max_{j \in \mathbb{N}(J)} c_j]^J$ . For the Berry et al. model, the non-compensatory treatment of income makes choosing “arbitrary” initial conditions more difficult. We test 10 initial conditions drawn uniformly from  $[0, 19]$ , where the 70<sup>th</sup> percentile of income is approximately 19,000



**Figure 4.1** Transaction versus equilibrium prices for 993 vehicle models for a single 1,000,000 sample set under the [Boyd and Mellman \(1980\)](#) model.

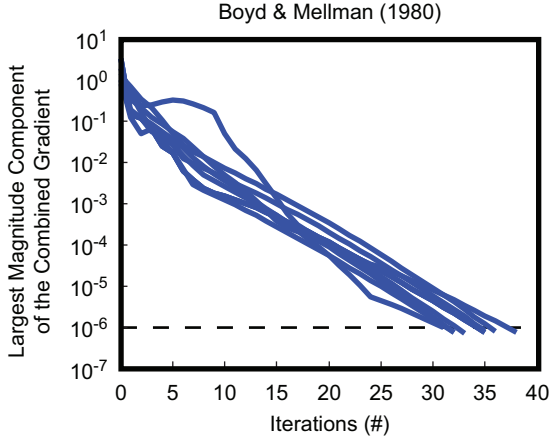


**Figure 4.2** Transaction versus equilibrium prices for 993 vehicle models for a single 1,000,000 sample set under the [Berry et al. \(1995\)](#) model.

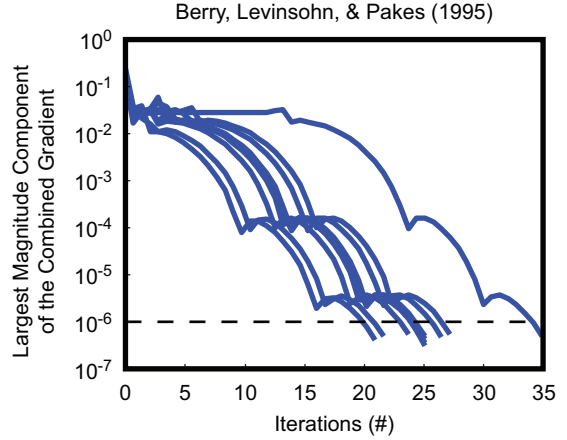
1983 USD. Using such initial conditions ensures that, loosely speaking, the upper 30 % of the sampled population can buy *any* vehicle at the initial prices, and does *not* preclude the existence of individuals in the sampled population that have an income too low to buy any of the vehicles at their initial prices. The combined gradient norm  $\|\tilde{\nabla}\hat{\pi}(\mathbf{p})\|_{\infty}$  consistently decreases to  $\sim 10^{-6}$  in less than forty fixed-point steps over these trials for various sample set sizes under both demand models. Furthermore, the computed equilibrium prices consistently generate negative definite profit Hessians for each firm. That is, the fixed-point iteration reliably computes local equilibrium prices, rather than simply simultaneously stationary prices.

Figs. 4.3 and 4.4 also illustrate quite different paths to simultaneous stationarity under the two models. Particularly, under the [Berry et al.](#) model,  $\|\tilde{\nabla}\hat{\pi}(\mathbf{p})\|_{\infty}$  need not be *monotonically* decreased over the iteration. Although this can be a feature of the trajectory of the combined gradient norm, the fixed-point step norm  $\|\zeta(\mathbf{p}) + \mathbf{c} - \mathbf{p}\|_{\infty}$  and the distance to equilibrium  $\|\mathbf{p} - \mathbf{p}^*\|_{\infty}$  may decrease monotonically.

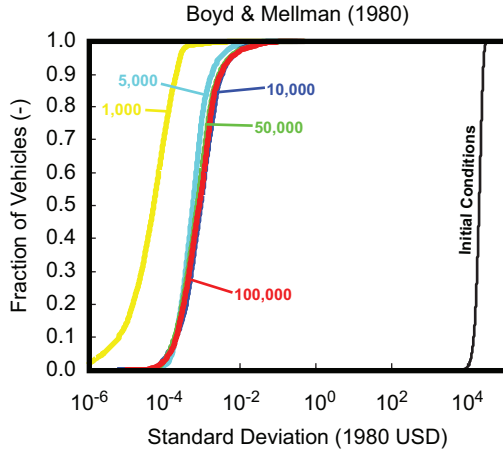
Figs. 4.5 and 4.6 illustrate the negligible variability in computed equilibrium prices driven by these “arbitrary” initial conditions. These figures plot the cumulative distribution of standard deviation in computed equilibrium prices over the entire set of model-year variants for the 10 initial conditions with a fixed sample set of various sizes (as well as the corresponding standard deviation in initial conditions). Under the [Boyd and Mellman](#) model all of the 5,298 model-year variants have standard deviations less than 1 1983 USD over these different initial conditions and sample set sizes. With 100,000 samples, 97% of these vehicles have standard deviations less than 0.01 1983



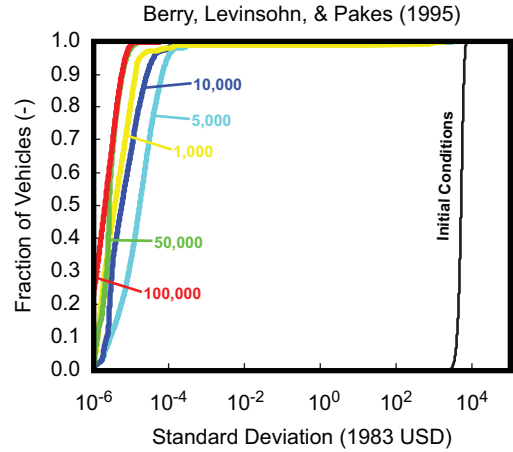
**Figure 4.3** Change in  $\|(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_\infty$  for 10 random initial conditions and a fixed sample set ( $S = 100,000$ ) under the [Boyd and Mellman \(1980\)](#) model.



**Figure 4.4** Change in  $\|(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_\infty$  for 10 random initial conditions and a fixed sample set ( $S = 100,000$ ) under the [Berry et al. \(1995\)](#) model.



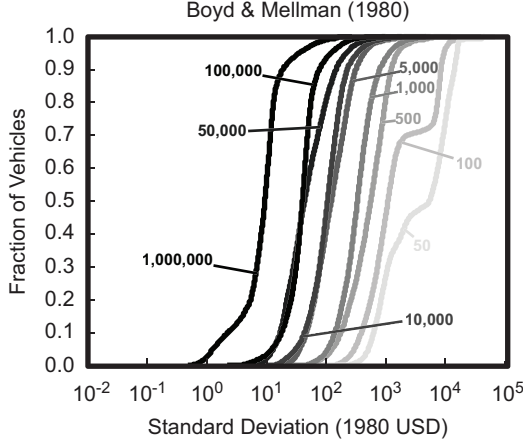
**Figure 4.5** Standard deviation in computed equilibrium prices over 10 random initial conditions for several sample set sizes (given in the numeric labels) under the [Boyd and Mellman \(1980\)](#) model.



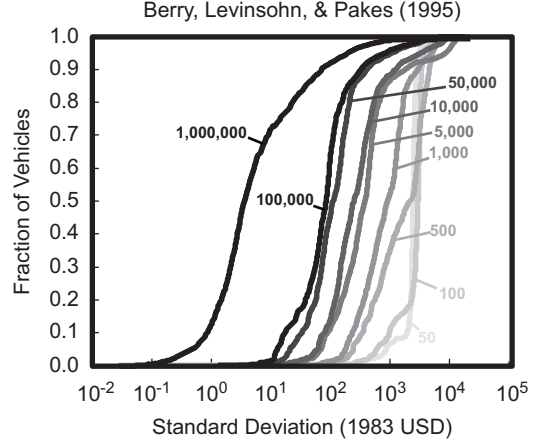
**Figure 4.6** Standard deviation in computed equilibrium prices over 10 random initial conditions for several sample set sizes (given in the numeric labels) under the [Berry et al. \(1995\)](#) model.

USD. Under the [Berry et al.](#) model, more than 99% of the 5,298 model-year variants have standard deviations less than 0.01 1983 USD. For the larger sample set sizes of 50,000 and 100,000 *all* computed equilibrium prices vary less than 0.01 1983 USD. Thus it appears that for any fixed sample set, regardless of size, computed equilibrium prices appear very stable over arbitrary initial conditions when using the fixed-point iteration.

Finally Figs. [4.7](#) and [4.8](#) demonstrate that computed equilibrium prices have a stronger dependence on the sample set used. We plot the cumulative distribution



**Figure 4.7** Cumulative distribution of standard deviation in computed equilibrium prices over 10-20 sample sets of several sizes (given in the numeric labels) under the [Boyd and Mellman \(1980\)](#) model.



**Figure 4.8** Cumulative distribution of standard deviation in computed equilibrium prices over 10-20 sample sets of several sizes (given in the numeric labels) under the [Berry et al. \(1995\)](#) model.

functions of standard deviation in computed equilibrium prices over different sample set sizes (using unit costs as the initial condition). As the Law of Large Numbers would suggest, increasing the sample size reliably decreases the standard deviation in computed equilibrium price for most vehicles. This holds under the [Boyd and Mellman](#) model for all sample set sizes, though becomes evident for the [Berry et al.](#) model only for sample sizes above 1,000. Small sample sizes ( $S = 50, 100, \text{ and } 500$ ) are problematic under the [Berry et al.](#) model because of the non-compensatory decisions engendered by the treatment of income. Even for very large sample sets ( $S = 1,000,000$ ), only 20-25% of vehicles have computed equilibrium prices varying *more* than 100 1980 USD under the two models. Thus precision in equilibrium price computations appears to be driven by the sample set, rather than the initial condition.

## 4.5 Method Comparison

In this section we compare the performance the four approaches to computing equilibrium prices identified in the introduction: CG-NM,  $\eta$ -NM,  $\zeta$ -NM, and  $\eta$ -FPI. To undertake this comparison we aggregate the characteristics of the 5,298 model-year variants to create a smaller, 993 model-year vehicle dataset. Specifically, we construct model-year vehicle characteristics by averaging the characteristics of the model-year variants of specific model-year vehicles. We do not currently have the detailed sales data required to construct a more preferable sales-weighted average of model-year

variant characteristics. Using this smaller dataset allows us to keep the computational times reasonable enough to quickly run numerous trials while maintaining a reasonably large and complex numerical problem. We do not see any significant reason beyond increased computational times that this comparison could not be made on the larger 5,298 vehicle problem, nor any reason to expect that the qualitative conclusions of the results described here would change.<sup>7</sup>

We only report results using Viswanath’s GMRES-Hookstep method (Viswanath, 2007, 2008). We have tested other variants of Newton’s method, finding many of them less robust and more time consuming for the example problems. In order to carry out our comparison on a “level field,” we have implemented these methods ourselves following the prescriptions in Dennis and Schnabel (1996), as well as many other more specific sources from the literature. While we have attempted to use the most efficient and robust known Newton methods, professionally developed Newton solvers are likely to have advantages that our implementations lack. However the same could be said of the fixed-point iteration which, after continued study, is likely to result in further refinements. Additionally, comparisons with matlab’s `fpsolve` have verified the efficiency of our implementations. However, there is no “professional” implementation of the GMRES-Hookstep. We have made every effort to make ours as professional as possible by following the prescriptions of Dennis and Schnabel (1996) for general trust region methods.

Other algorithmic details are as follows.  $\|(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_{\infty} \leq 10^{-6}$  is the termination condition for all methods, and the second order conditions are always checked at computed equilibria. GMRES in CG-NM uses a problem-specific preconditioner discussed in Appendix C. Neither  $\eta$ -NM or  $\zeta$ -NM require preconditioned GMRES. We require that GMRES always solve the (unpreconditioned) Newton system to a relatively tight relative residual of  $10^{-6}$  within 50 steps without restarting. Loosening this relative residual tolerance<sup>8</sup> or including the adaptive tolerances recommended by Eisenstat and Walker (1996) may incrementally improve computational speed. These inclusions should not make a significant impact on the  $\eta$ -NM or  $\zeta$ -NM methods simply because GMRES is already very fast on these methods: generally less than ten GMRES steps are required to achieve convergence to an inexact Newton step. For CG-NM, these inclusions may have a more pronounced improvement.

---

<sup>7</sup>If we were using direct methods to solve the Newton system, the order of magnitude increase in problem size may in fact be good reason to expect degradation in the speed and quality of Newton’s method.

<sup>8</sup>Viswanath (2008) uses a relative residual tolerance of only  $10^{-2}$ .

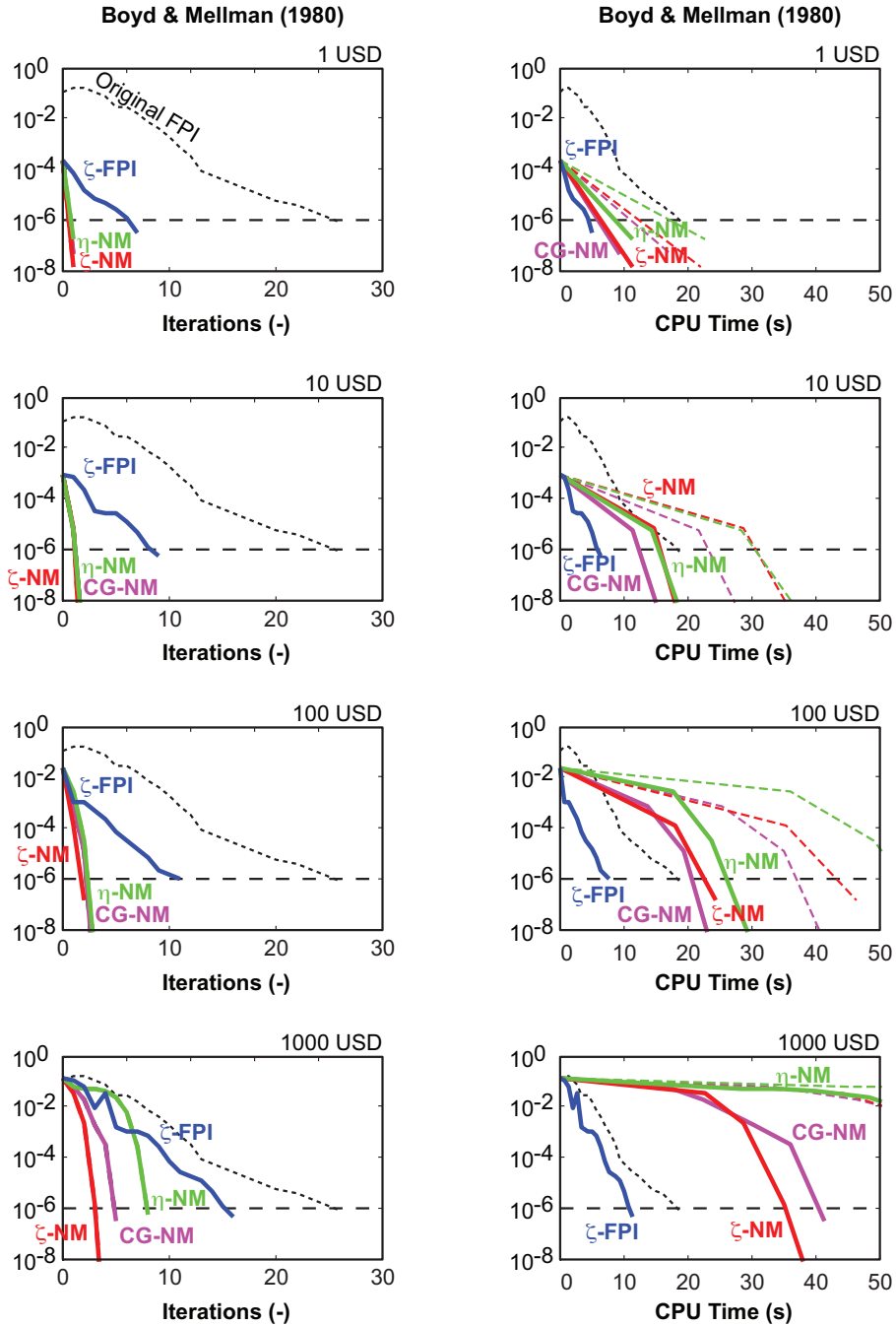
## Starting Near Equilibrium

To test that Newton’s method performs well when started near equilibrium prices (as predicted by theory; e.g., see [Ortega and Rheinboldt \(1970\)](#); [Dennis and Schnabel \(1996\)](#)), we performed the following test. We first compute equilibrium prices  $\mathbf{p}^*$  using the fixed-point iteration for a single 1,000 sample set. We then generate a set of  $T = 10$  initial conditions  $\mathbf{p}^{(t)} = \mathbf{p}^* + \kappa\boldsymbol{\nu}^{(t)}$ ,  $t \in \mathbb{N}(T)$ , by drawing  $\boldsymbol{\nu}^{(t)}$  from a uniform distribution on  $[-\mathbf{1}, \mathbf{1}]$ . The multiplier  $\kappa$  given the values 1, 10, 100, and 1000 USD. One can think of these  $T$  initial conditions as guesses of equilibrium prices correct to within  $\kappa$  USD.

Table 4.3 contains a summary of results for this experiment under the [Boyd and Mellman \(1980\)](#) model. All methods appear to be a reliable way to solve for equilibrium prices. Except for a single trial with the CG-NM formulation and a 1,000 USD perturbation, all trials converged to simultaneously stationary prices satisfying the second order conditions. All methods converge rapidly for small perturbations ( $< 100$  USD), with the Newton method approaches taking only 1 and 2 steps for 1 and 10 USD perturbations, respectively. The  $\boldsymbol{\eta}$ -NM method converges in very few iterations even with perturbations of 1,000. Another feature to note is the speed of the  $\boldsymbol{\zeta}$ -FPI method. Generally speaking, the approaches based on Newton’s method take 250-770% of the time required by the  $\boldsymbol{\zeta}$ -FPI to converge despite taking 30-50% fewer iterations. Particularly, even though CG-NM,  $\boldsymbol{\eta}$ -NM, and  $\boldsymbol{\zeta}$ -NM converge in a single step under a 1 USD perturbation, they take at least twice as long to do so as the fixed-point iteration.

Fig. 4.9 illustrates a typical performance by the various approaches we consider under the [Boyd and Mellman](#) model. These plots clearly illustrate the rapid convergence of Newton’s method applied to any formulation, in terms of iterations. However when considering computation (CPU) time, we again observe that, for this example, a single step of Newton’s method can take longer than it takes  $\boldsymbol{\zeta}$ -FPI to converge. For 100 and 1,000 USD perturbations, using Newton’s method took longer than the original  $\boldsymbol{\zeta}$ -FPI started at unit costs and used to compute equilibrium prices, regardless of the formulation to which Newton’s method is applied. One can also observe in Fig. 4.9 the significant speed-up gained by using the analytical Jacobian for matrix multiplications in **GMRES**. Using analytical and directional finite difference Jacobians result in the same convergence curves in terms of iterations, but using directional finite differences can double the time required by computations.

Table 4.4 details a similar comparison under the [Berry et al. \(1995\)](#) model. Again, all methods are fairly robust near equilibrium, with only three exceptions. In one



**Figure 4.9** Typical convergence curves for perturbation trials under the [Boyd and Mellman \(1980\)](#) model. The fixed-point iteration and the GMRES-Hookstep Newton’s method are starting from random points in  $\mathbf{p}^* + \kappa[-\mathbf{1}, \mathbf{1}]$  USD, where  $\mathbf{p}^*$  is a vector of equilibrium prices computed using  $\zeta$ -FPI (convergence curve given as dotted black line). Results for a fixed sample set ( $S = 1,000$ ) under the [Boyd and Mellman \(1980\)](#) model. Solid colored lines denote convergence curves for Newton method implementations using the analytical Jacobian, and the colored dashed lines denote convergence curves for Newton method implementations using directional finite differences.

**Table 4.3** Summary of results for ten perturbation trials under the [Boyd and Mellman \(1980\)](#) model for a fixed set of 1,000 samples. Both iterations and CPU time are given in the form “minimum/median/maximum,” each taken over the ten trials.

|          |            |     | <a href="#">Boyd and Mellman (1980)</a> |            |             |              |
|----------|------------|-----|---|------------|-------------|--------------|
|          |            |     | CG-NM                                   | $\eta$ -NM | $\zeta$ -NM | $\zeta$ -FPI |
| 1 USD    | Iterations | (#) | 1/1/1                                   | 1/1/1      | 1/1/1       | 5/7/8        |
|          | CPU Time   | (s) | 15/15/16                                | 18/18/18   | 18/18/18    | 4/5/6        |
| 10 USD   | Iterations | (#) | 2/2/2                                   | 2/2/2      | 2/2/2       | 8/11/11      |
|          | CPU Time   | (s) | 23/23/24                                | 27/27/28   | 26/26/27    | 6/8/9        |
| 100 USD  | Iterations | (#) | 2/3/4                                   | 2/3/3      | 2/3/4       | 12/14/15     |
|          | CPU Time   | (s) | 25/31/36                                | 30/33/36   | 30/35/41    | 9/10/11      |
| 1000 USD | Iterations | (#) | 4/6/14 <sup>(a)</sup>                   | 3/4/5      | 6/7/10      | 15/17/19     |
|          | CPU Time   | (s) | 39/47/94 <sup>(a)</sup>                 | 41/48/53   | 57/68/82    | 11/12/14     |

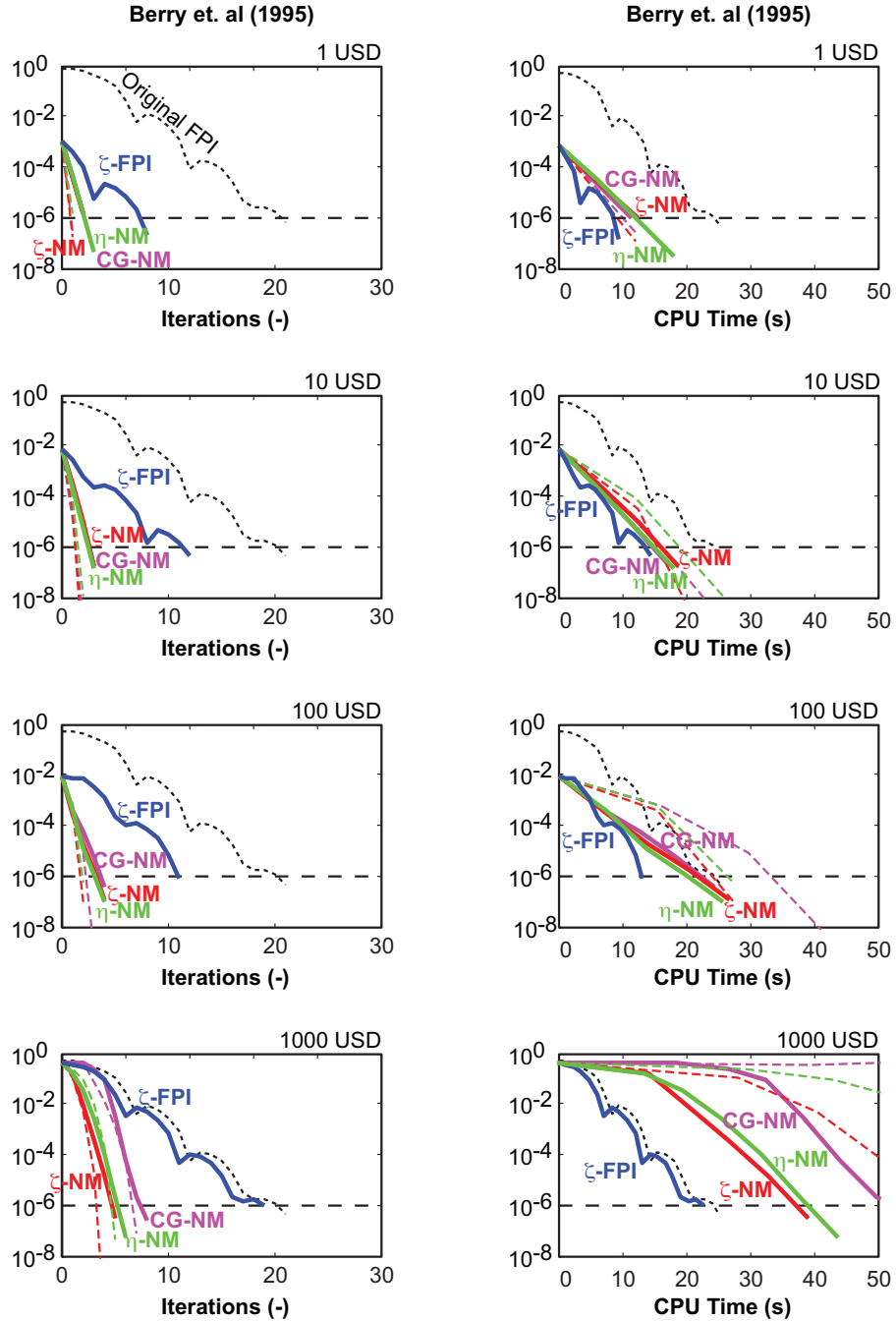
(a) The results of two trials failed to satisfy the second order conditions.

trial,  $\zeta$ -NM failed to compute simultaneously stationary prices. In two separate trials, CG-NM failed to compute simultaneously stationary prices and compute prices satisfying the second order condition. Again, Newton’s method across formulations takes 30-60% of the iterations required of the  $\zeta$ -FPI but 150-360% of the CPU time. The  $\zeta$ -FPI has the significant advantage of requiring, on average, only  $\sim 1.2$  seconds per iteration regardless of  $\kappa$  while Newton’s method requires 6 seconds per iteration for small  $\kappa$  and 7 for large  $\kappa$ . This increase in average time per iteration is, of course, related to more relatively more expensive hooksteps becoming required as the initial prices deviate from equilibrium.

Fig. 4.10 illustrates a typical performance by the approaches we consider under the [Berry et al.](#) model. Again, these plots clearly illustrate the rapid convergence of Newton’s method applied to any formulation, in terms of iterations. For 100 and 1,000 USD perturbations, using Newton’s method again took longer than the original  $\zeta$ -FPI started at unit costs. Though one can still observe in Fig. 4.10 a CPU time benefit to using the analytical Jacobian for matrix multiplications in **GMRES**, the advantage is somewhat less pronounced under the [Berry et al. \(1995\)](#) model. Interestingly, it appears that using directional finite differences increases the convergence rate in terms of iterations, but still ultimately increases overall computation times.

Under both models, the prices computed by the fixed-point iteration and the variants of Newton’s method compare quite well to the equilibrium prices used to generate the perturbed initial conditions. Using any convergent method, roughly 97%





**Figure 4.10** Typical convergence curves for perturbation trials under the [Berry et al. \(1995\)](#) model. The fixed-point iteration and the GMRES-Hookstep Newton’s method are starting from random points in  $\mathbf{p}^* + \kappa[-\mathbf{1}, \mathbf{1}]$  USD, where  $\mathbf{p}^*$  is a vector of equilibrium prices computed using  $\zeta$ -FPI (convergence curve given as dotted black line). Results for a fixed sample set ( $S = 1,000$ ) under the [Berry et al. \(1995\)](#) model. Solid colored lines denote convergence curves for Newton method implementations using the analytical Jacobian, and the colored dashed lines denote convergence curves for Newton method implementations using directional finite differences.

**Table 4.4** Summary of results for ten perturbation trials under the [Berry et al. \(1995\)](#) model for a fixed set of 1,000 samples. Both iterations and CPU time are given in the form “minimum/median/maximum,” each taken over the ten trials.

|          |            |     | <a href="#">Berry et al. (1995)</a> |            |                          |              |
|----------|------------|-----|-------------------------------------|------------|--------------------------|--------------|
|          |            |     | CG-NM                               | $\eta$ -NM | $\zeta$ -NM              | $\zeta$ -FPI |
| 1 USD    | Iterations | (#) | 2/2/3                               | 2/2/3      | 2/2/3                    | 8/9/10       |
|          | CPU Time   | (s) | 18/18/24                            | 20/20/26   | 19/19/25                 | 10/11/12     |
| 10 USD   | Iterations | (#) | 3/3/3                               | 3/3/3      | 3/3/3                    | 10/12/13     |
|          | CPU Time   | (s) | 24/24/24                            | 24/26/26   | 25/25/27                 | 12/14/15     |
| 100 USD  | Iterations | (#) | 4/4/5                               | 3/4/4      | 3/4/5                    | 12/14/17     |
|          | CPU Time   | (s) | 31/31/37                            | 27/34/34   | 27/33/39                 | 14/17/20     |
| 1000 USD | Iterations | (#) | 8/10/50 <sup>(a)</sup>              | 5/5/6      | 5/6/26 <sup>(b)</sup>    | 17/18/22     |
|          | CPU Time   | (s) | 58/79/320 <sup>(a)</sup>            | 46/46/52   | 45/52/248 <sup>(b)</sup> | 20/21/26     |

(a) CG-NM failed to compute simultaneously stationary prices in one trial, and failed to compute prices satisfying the second order conditions in another.

(b)  $\zeta$ -NM failed to compute simultaneously stationary prices in one trial.

of vehicles have prices deviating from the originally computed equilibrium prices by less than 1 USD (in the appropriately deflated units).

### Starting at Unit Costs

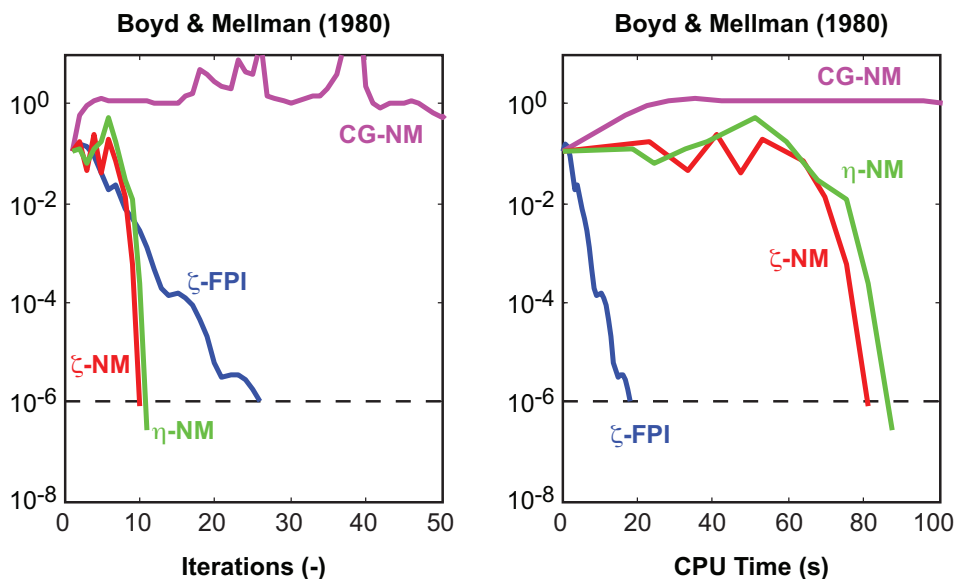
Unit costs give a “canonical” starting point in the absence of a reasonable guess about equilibrium prices. [Table 4.5](#) details results starting from unit costs, with typical convergence curves provided in [Figs. 4.11](#) and [4.12](#). The GMRES-Hookstep method generally failed to converge for the CG-NM problem under both demand models, and thus is not included in [Table 4.5](#). Both the  $\eta$ -NM and  $\zeta$ -NM problems converge reliably, as does the  $\zeta$ -FPI. Again, despite a rapid convergence rate in terms of iterations, the relative computational intensity of taking Newton steps makes the  $\zeta$ -FPI the fastest method. For the example illustrated in [Fig. 4.11](#), the  $\eta$ -NM method requires only about 30% of the iterations of the  $\zeta$ -FPI but 440% of the actual CPU time. For the example illustrated in [Fig. 4.12](#), the  $\zeta$ -NM method requires only about 25% of the iterations of the  $\zeta$ -FPI but 200% of the actual CPU time.

**Table 4.5** Results of price equilibrium computations starting at unit costs under both demand models for ten 1,000-sample sets. Iterations and CPU time are again listed as “minimum/median/maximum,” where we only include successful trials; that is, trials in which a simultaneously stationary point was computed. The number of successful trials is also given below. All successful trials resulted in prices satisfying the second order conditions.

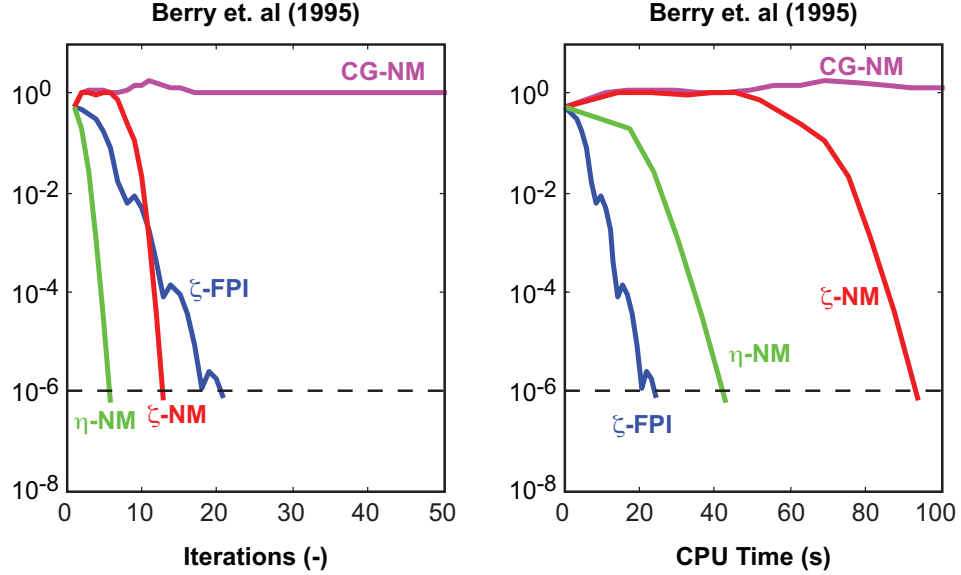
|            |       | Boyd and Mellman (1980) |             |              |
|------------|-------|-------------------------|-------------|--------------|
|            |       | $\eta$ -NM              | $\zeta$ -NM | $\zeta$ -FPI |
| Iterations | (#)   | 8/10/15                 | 10/15/20    | 23/39/70     |
| CPU Time   | (s)   | 71/87/123               | 80/118/147  | 17/29/53     |
| Successful | (#/#) | 7/10                    | 6/10        | 10/10        |

|            |       | Berry et al. (1995) |             |              |
|------------|-------|---------------------|-------------|--------------|
|            |       | $\eta$ -NM          | $\zeta$ -NM | $\zeta$ -FPI |
| Iterations | (#)   | 6/6/9               | 8/9/11      | 18/19/23     |
| CPU Time   | (s)   | 50/52/70            | 64/73/90    | 23/24/29     |
| Successful | (#/#) | 10/10               | 9/10        | 10/10        |



**Figure 4.11** Convergence curves for the fixed-point iteration and variants of Newton’s method started at unit costs for a fixed 1,000 sample set under the [Boyd and Mellman \(1980\)](#) model. Newton’s method applied to the preconditioned combined gradient does not converge.



**Figure 4.12** Convergence curves for the fixed-point iteration and variants of Newton’s method started at unit costs for a fixed 1,000 sample set under the [Berry et al. \(1995\)](#) model. Newton’s method applied to the preconditioned combined gradient does not converge.

## 4.6 Conclusions

In this article we have described several methods for computing equilibrium prices under a nearly arbitrary Mixed Logit model. One approach naively applies Newton’s method to the simultaneous stationarity condition. Two other approaches apply Newton’s method to distinct fixed-point characterizations derived from the simultaneous stationarity condition. Finally, one of these fixed-point characterizations, novel to this work, is used in a fixed-point iteration. We have undertaken numerical comparison of these approaches using two models of the 2005 new vehicle market and a state-of-the-art Newton’s method for large-scale systems, Viswanath’s GMRES-Hookstep ([Viswanath, 2007, 2008](#)). These examples have demonstrated the ability to reliably and efficiently compute equilibrium prices in markets with many products and a large degree of product and firm heterogeneity under Mixed Logit demand specifications using the fixed-point characterizations of equilibrium prices. While the GMRES-Hookstep applied to either of the fixed-point characterizations appears to be an efficient and reliable approach for computing equilibrium prices, the fixed-point iteration was the most reliable and efficient method of computing equilibrium prices for our example problems.

We thus view the fixed-point iteration proposed in this article as a powerful alternative to Newton’s method when economists, marketers, operations researchers,

and engineers need to compute equilibrium prices in differentiated product market models. The fixed-point iteration appears to be an efficient computational tool even when counterfactual assumptions might radically alter the structure of equilibrium prices. The added reliability and efficiency of the fixed-point iteration should allow for more developed counterfactual analyses exploring model outcomes over a variety of counterfactual scenarios, including extreme conditions. This “reliable efficiency” may also prove useful when equilibrium pricing is a component of optimization approaches to determining governmental or firm strategy.

# Chapter 5

## Regulated Equilibrium Prices Under Mixed Logit Models

In this chapter, we study regulated Bertrand competition under Mixed Logit models of demand, restricting attention to several regulatory cost forms relevant for the automotive industry. First, Section 5.1 presents an extended framework that incorporates regulatory costs. Next in Section 5.2 we describe several regulatory policies that fit under this framework. Using a generalized form of expected regulatory costs, in Section 5.3 we derive an extension of the  $\zeta$  fixed-point characterization, valid even when expected regulatory costs are not differentiable. We also illustrate how these forms can shed some light on incentives and “shadow taxes” under certain regulatory policies. In Section 5.4 we propose a numerical approach for the computation of regulated equilibrium prices motivated by the success of the  $\zeta$ -FPI. For differentiable regulatory costs, the extended  $\zeta$ -FPI suffices, while for standards with non differentiable regulatory costs  $\zeta$ -FPI must be combined with direct solution of the first-order non-smooth necessary conditions.

### 5.1 Additions to the Framework

In this section we describe an extension to the framework discussed in Chapter 2 that includes regulatory costs.

#### 5.1.1 Regulatory Costs and Heuristic Expected Profits

We assume that firms face regulatory costs prescribed by a firm-specific penalty level  $r_f \geq 0$  and a function  $c_f^R : \mathbb{Z}_+^J \times \mathcal{Y}^J \rightarrow \mathbb{R}$  that *may* depend on all quantities sold (and

hence produced) as well as the product characteristics of all products offered. This form is a generalization of several regulatory policies proposed for the automotive industry, discussed below. Note also that we allow “costs” to take on negative values, as occurs with subsidies or rebates. With this form, the total (random) regulatory cost imposed on firm  $f$  is then given by  $r_f c_f^R(\mathbf{Q}(\mathbf{Y}, \mathbf{p}), \mathbf{Y})$ .

Without production capacity constraints that limit a firm’s ability to meet any demands that arise during the purchase period and assuming that unit and fixed costs remain independent of the specific quantity demanded (and hence prices), firm  $f$ ’s profits as a function of prices are given by the random variable  $\Pi_f(\mathbf{p}) = \mathbf{Q}_f(\mathbf{Y}, \mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f^U) - c_f^F - r_f c_f^R(\mathbf{Q}_f(\mathbf{Y}, \mathbf{p}), \mathbf{Y})$ . The expected regulatory costs complicate the evaluation of profits because they can, in principle and in practice, be a nonlinear function of demands. Computing with actual expected regulatory costs  $\mathbb{E}[c_f^R(\mathbf{Q}(\mathbf{Y}, \mathbf{p}), \mathbf{Y})]$ , and its price derivatives, requires approximating sums over the very large space of realizable demands, and we are not currently aware of tractable ways to do this in the context of equilibrium pricing problems.

Existing analyses of regulated Bertrand competition under a RUM implicitly assume that

$$\mathbb{E}[c_f^R(\mathbf{Q}(\mathbf{Y}, \mathbf{p}), \mathbf{Y})] = c_f^R(\mathbb{E}[\mathbf{Q}(\mathbf{Y}, \mathbf{p})], \mathbf{Y})$$

although may not be true for the policy being studied (e.g. [Goldberg \(1998\)](#)). In this work we follow the existing literature in using  $c_f^R(\mathbb{E}[\mathbf{Q}(\mathbf{Y}, \mathbf{p})], \mathbf{Y})$  although we *recognize this as a heuristic*.

**Assumption 5.1.1.** *Firms choose prices to maximize  $I\hat{\pi}_f^R(\mathbf{p}) - c_f^F$  where  $\hat{\pi}_f^R(\mathbf{p}) = \hat{\pi}_f(\mathbf{p}) - r_f \hat{c}_f^R(\mathbf{p})$  and  $\hat{c}_f^R(\mathbb{E}[\mathbf{Q}(\mathbf{Y}, \mathbf{p})], \mathbf{Y}) = I^{-1} c_f^R(\mathbb{E}[\mathbf{Q}(\mathbf{Y}, \mathbf{p})], \mathbf{Y})$ .*

The computational burden of determining actual expected regulatory costs may justify the use of the heuristic expected regulatory costs  $\hat{c}_f^R(\mathbf{Y}, \mathbf{p})$  in market models. Specifically, firms may not use the actual expected regulatory costs in their decision making because this is too difficult to evaluate while heuristic is relatively easy to evaluate. On the other hand, much of this work concerns overcoming one key technical shortcoming of the heuristic expected regulatory costs: that they are not always differentiable functions of  $\mathbf{p}$ . The true expected regulatory costs are (continuously) differentiable in  $\mathbf{p}$  so long as  $\mathbb{P}(\mathbf{Q}(\mathbf{Y}, \mathbf{p}) = \mathbf{q})$  is a (continuously) differentiable function of  $\mathbf{p}$ , *regardless* of nonlinearities in  $c_f^R$ .

In all of our examples, the heuristic expected regulatory costs can be written in a specific form that should hold for any regulatory policy that scales regulatory penalties (or revenues) with total demand to the firm.

**Assumption 5.1.2.** *There exists a map  $\boldsymbol{\varrho}_f : \mathbb{R}^J \rightarrow \mathbb{R}^{J_f}$  such that  $\hat{c}_f^R(\mathbf{p}) = \mathbf{P}_f(\mathbf{p})^\top \boldsymbol{\varrho}_f(\mathbf{p})$  for all  $\mathbf{p} \in \mathbb{R}^J$ .*

Assumption 5.1.2 implies that regulated profits can be written

$$\hat{\pi}_f^R(\mathbf{p}) = \mathbf{P}_f(\mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f - r_f \boldsymbol{\varrho}_f(\mathbf{p})).$$

Some policies have non-differentiable  $\boldsymbol{\varrho}_f$ . In this work we are only concerned with non-differentiable policies satisfying the following assumption.

**Assumption 5.1.3 (Standards).** *There exists a continuously differentiable function  $\hat{\varrho}_f : [\mathbf{0}, \varsigma_* \mathbf{1}] \rightarrow \mathbb{R}$  such that  $\boldsymbol{\varrho}_f(\mathbf{p}) = \max\{0, \hat{\varrho}_f(\mathbf{p})\} \mathbf{1}$ .*

**Definition 5.1.1.** *Policies satisfying Assumption 5.1.3 are called standards.  $\hat{\varrho}_f$  is called the standard function. We say that the standard is violated at  $\mathbf{p}$  if  $\hat{\varrho}_f(\mathbf{p}) > 0$ , satisfied at  $\mathbf{p}$  if  $\hat{\varrho}_f(\mathbf{p}) < 0$ , and exactly satisfied at  $\mathbf{p}$  if  $\hat{\varrho}_f(\mathbf{p}) = 0$ . Prices satisfying  $\hat{\varrho}_f(\mathbf{p}) = 0$  will also be called standard-bound.*

## 5.2 Several Regulatory Policies

We now introduce several regulatory policies debated for the automotive industry.

Throughout,  $\kappa : \mathcal{Y} \rightarrow \mathbb{R}$  denotes an arbitrary function of vehicle characteristics; e.g., fuel economy, fuel consumption, or CO<sub>2</sub> emissions intensity. This function may be a projection – fuel economy itself may be a characteristic – or may be a proper function of characteristics. For example, we may want  $\kappa$  to be fuel economy while fuel cost, defined as fuel price times fuel consumption, is a characteristic. Let also  $\boldsymbol{\kappa} : \mathcal{Y}^J \rightarrow \mathbb{R}^J$  be defined componentwise by  $\kappa_j(\mathbf{Y}) = \kappa_j(\mathbf{y}_j) = \kappa(\mathbf{y}_j)$  and, for each  $f$ , let  $\boldsymbol{\kappa}_f(\mathbf{Y}) \equiv \boldsymbol{\kappa}_f(\mathbf{Y}_f)$  denote the firm-specific sub-vectors of  $\boldsymbol{\kappa}(\mathbf{Y})$ . For any vectors  $\mathbf{x}, \mathbf{w} \in \mathbb{R}^N$ ,  $\mathbf{w} \geq \mathbf{0}$ , we let  $\mathbf{a}^A(\mathbf{x}, \mathbf{w})$  denote the  $\mathbf{w}$ -weighted arithmetic average of the values in the vector  $\mathbf{x}$  and  $\mathbf{a}^H(\mathbf{x}, \mathbf{w})$  denotes the  $\mathbf{w}$ -weighted harmonic average of the values in the vector  $\mathbf{x}$ . That is,

$$\mathbf{a}^A(\mathbf{x}, \mathbf{w}) = \frac{\sum_{n=1}^N w_n x_n}{\sum_{n=1}^N w_n} \quad \text{and} \quad \mathbf{a}^H(\mathbf{x}, \mathbf{w}) = \frac{\sum_{n=1}^N w_n}{\sum_{n=1}^N w_n / x_n}.$$

### 5.2.1 Fuel Taxes.

Fuel taxes are one prominent regulatory policy option that does not directly explicitly regulate the firm at all, but has an effect on pricing decisions (Verboven, 2002). These



taxes often enter the differentiated product market model through operating cost characteristics in the demand model, and thus can be numerically treated using techniques such as those described in Chapter 4.

### 5.2.2 CO<sub>2</sub> taxes on firms.

Michalek et al. (2004) analyze a tax levied on firms proportional to the expected lifetime CO<sub>2</sub> emitted by driving vehicles they sell. Letting  $\boldsymbol{\kappa}_f(\mathbf{Y}_f)$  denote the vector of expected lifetime CO<sub>2</sub> emissions for each vehicle as a function of product characteristics, firm  $f$  faces regulatory costs  $c_f^R(\mathbf{q}, \mathbf{Y}) = \mathbf{q}_f^\top \boldsymbol{\kappa}_f(\mathbf{Y}_f)$ . Fixing  $\mathbf{Y}_f$  and writing  $\boldsymbol{\kappa}_f = \boldsymbol{\kappa}_f(\mathbf{Y}_f)$ , this formula yields the heuristic expected regulatory costs  $\hat{c}_f^R(\mathbf{p}) = \mathbf{P}_f(\mathbf{p})^\top \boldsymbol{\kappa}_f$ , which in this case (and this case only) are also the actual expected regulatory costs. Thus  $\boldsymbol{\varrho}_f(\mathbf{p}) = \boldsymbol{\kappa}_f$ , a constant. More generally, for any fixed vector  $\mathbf{t}_f \in \mathbb{R}^{J_f}$  we call the policy with  $\boldsymbol{\varrho}_f \equiv \mathbf{t}_f$  a *flat tax/subsidy* policy.

The effects of this policy might be regarded as equivalent to the impact of a cap-and-trade policy that makes *firms* accountable for the lifetime CO<sub>2</sub> emissions of the vehicles they sell. Firm  $f$  chooses to purchase an allowance of  $a_f$  units of emission from the permit market at a price  $\rho$ , which we assume the firm alone cannot control (it's set on a world market). That is, we assume that competition for permits is perfect, in the sense that no individual firm can change the permit price by changing their product prices. Firm  $f$  is charged  $r_f \geq 0$  USD for every unit of emission above  $a_f$  it is accountable for. This makes firm  $f$ 's regulated profits

$$\hat{\pi}_f^R(\mathbf{p}, a_f) = \hat{\pi}_f(\mathbf{p}) - \rho a_f - r_f (\mathbf{P}_f(\mathbf{p})^\top \boldsymbol{\kappa}_f - a_f) = \hat{\pi}_f(\mathbf{p}) - (\rho - r_f) a_f - r_f (\mathbf{P}_f(\mathbf{p})^\top \boldsymbol{\kappa}_f).$$

Note that  $(\nabla_{\mathbf{p}}^p \hat{\pi}_f^R)(\mathbf{p}, a_f) = \nabla_{\mathbf{p}}^p [\hat{\pi}_f(\mathbf{p}) - r_f (\mathbf{P}_f(\mathbf{p})^\top \boldsymbol{\kappa}_f)]$ , and thus the price gradient is the same as under a CO<sub>2</sub> tax levied on firms (getting all the quantities right). In particular, profit-stationary prices are independent of the allowance procured by the firm and are those of a model with a CO<sub>2</sub> tax of  $r_f$ .

We must remark that accounting for driving CO<sub>2</sub> emissions through the manufacturer is not an effective way to implement a cap-and-trade program aimed at capping the CO<sub>2</sub> emissions from driving. Funneling the cap policy through automotive firms would be flawed for two reasons. First, the “caps” cannot be respected because only *expected* lifetime emissions are accounted for during the sale year of the vehicle. There is no control over the actual value or temporal distribution of these emissions, antithetical to the concept of an effective cap. Second, drivers of the vehicles should be incentivized under the cap-and-trade program to reduce their emissions as well

as modify their vehicle purchases. This might be effectively accomplished through the fuel producers if it can be assumed that vehicles have identical emissions factors. Though complicated and politically challenging, an effective cap-and-trade program for the transportation industry must target individual drivers. CO<sub>2</sub> taxes as described above are a better model of the impacts of cap-and-trade on the CO<sub>2</sub> burden of vehicle production.

### 5.2.3 Corporate Average Fuel Economy (CAFE) Standards.

The Corporate Average Fuel Economy (CAFE) standards are the primary regulatory policy active in the U.S. automotive industry (NHTSA, 2008b). Let  $\kappa : \mathcal{Y} \rightarrow \mathbb{R}_+$  define a vehicle’s fuel economy from its characteristic vector, and  $\kappa_f^S$  be a fixed, firm-specific standard level.<sup>1</sup> A basic CAFE standard has the penalty

$$c_f^R(\mathbf{q}, \mathbf{Y}) = c_f^R(\mathbf{q}_f, \mathbf{Y}_f) = (\mathbf{q}_f^\top \mathbf{1}) \max \{ 0, \kappa_f^S - \mathbf{a}^H(\boldsymbol{\kappa}_f(\mathbf{Y}), \mathbf{q}_f) \}.$$

Thus, setting  $\boldsymbol{\kappa}_f = \boldsymbol{\kappa}_f(\mathbf{Y}_f)$ , this basic CAFE standard has heuristic expected regulatory costs

$$\hat{c}_f^R(\mathbf{p}) = (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \max \{ 0, \kappa_f^S - \bar{\kappa}_f^H(\mathbf{p}) \} \quad \text{where} \quad \bar{\kappa}_f^H(\mathbf{p}) = \mathbf{a}^H(\boldsymbol{\kappa}_f, \mathbf{P}_f(\mathbf{p}))$$

These can be written in a form satisfying Assumption 5.1.2 by setting  $\boldsymbol{\varrho}_f(\mathbf{p}) = \max\{0, \kappa_f^S - \bar{\kappa}_f^H(\mathbf{p})\}\mathbf{1}$ . We note that this  $\boldsymbol{\varrho}_f$  is *not* differentiable for any  $\mathbf{p}$  such that  $\bar{\kappa}_f^H(\mathbf{p}) = \kappa_f^S$ .

The actual CAFE standards are more complicated than this simple model, in primarily three ways. First is that there are different standards for different “classes” of products firms offer. Particularly, there are independent standards for “domestic” cars,<sup>2</sup> “imported” cars, and “light-duty” trucks.<sup>3</sup> Though we do not provide an explicit description of the extension to multi-class standards, most of our framework applies in the more general setting. Second, the existing crediting system is a persistent complication. Firms can receive “credits” for offering and selling alternative fuel vehicles, though the fuel consumption and emissions benefits of these alternative fuel vehicles need not actually be accrued in the fleet (NRC, 2002). Firms can also offset

<sup>1</sup>All firms currently share the same standard level under the CAFE standards. This will change in 2011, the year attribute-based standards begin.

<sup>2</sup>A car is domestically produced if “if at least 75 percent of the cost to the manufacturer is attributable to value added in the United States or Canada” (NHTSA, 2008d, 32904(b)(2)).

<sup>3</sup>“Heavy-duty” trucks having a Gross Vehicle Weight Rating over 8,500 pounds have historically been excluded from the CAFE standards. This weight rating limit has been raised to 10,000 pounds.

CAFE deficits in one year with CAFE surpluses within the following three years. Similarly, firms can credit CAFE surpluses in the current year towards CAFE deficits in following years. [Jacobsen \(2006\)](#) considers this temporal crediting system by simply using repeated Bertrand competition models with history-dependent CAFE standards reflecting past surpluses or deficits. Most generally, treating the temporal crediting system should require a pricing analysis where firms decide on pricing with foresight regarding future markets. A third element of complication, the use of standard levels that depend on the characteristics of vehicles, is our next item of discussion.

### Attribute-based CAFE

Based on an argument that the existing CAFE standards discriminate against the American automakers ([NRC, 2002](#); [Levin, 2007b](#)), the new CAFE standards passed in 2007 include flexible, firm-specific standards that depend on the mix of vehicles sold by a firm. Essentially, this is meant to bolster American firms which, although selling more larger, less fuel-efficient vehicles than its foreign rivals, is generally perceived to have more efficient large vehicles than the foreign competition. This has not yet been a prominent part of the active CAFE standards, having only been enacted by the NHTSA for the light-duty truck class as an option for 2008-2011 model years ([NHTSA, 2006](#)) and has just been introduced as a mandatory component of the national standards for the 2011-2015 model years [NHTSA \(2008a\)](#).

The flexible firm-specific standard is based on a continuous target fuel economy function defined over the vector of vehicle characteristics. Specifically, let  $\kappa^T : \mathcal{Y} \rightarrow \mathbb{R}_+$  define a fuel economy target as a function of vehicle characteristics. Letting  $\boldsymbol{\kappa}_f^T = \boldsymbol{\kappa}_f^T(\mathbf{Y}_f)$  be the vector of vehicle-specific targets for those vehicles offered by firm  $f$ , this firm is assigned an attribute-based CAFE standard of  $\mathbf{a}^H(\boldsymbol{\kappa}_f^T, \mathbf{q}_f)$  when demands for firm  $f$ 's vehicles are given by  $\mathbf{q}_f$ . Firm  $f$ 's basic regulatory costs are

$$c_f^R(\mathbf{q}_f, \mathbf{Y}_f) = (\mathbf{q}_f^\top \mathbf{1}) \max\{ 0, \mathbf{a}^H(\boldsymbol{\kappa}_f^T, \mathbf{q}_f) - \mathbf{a}^H(\boldsymbol{\kappa}_f, \mathbf{q}_f) \},$$

with heuristic expected regulatory costs given by

$$\hat{c}_f^R(\mathbf{p}) = (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \max\{ 0, \kappa_f^S(\mathbf{p}) - \bar{\kappa}_f^H(\mathbf{p}) \}.$$

where  $\kappa_f^S(\mathbf{p}) = \mathbf{a}^H(\boldsymbol{\kappa}_f^T, \mathbf{P}_f(\mathbf{p}))$  is the firm-specific (heuristic) standard level as a function of prices. NHTSA has chosen the target fuel economy,  $\kappa^T$ , to be generated by

**Table 5.1** Parameters proposed by NHTSA for attribute-based CAFE standards in 2011-2015. See Eqn. (5.1)

|     | Cars |      |      |      |      | Light Trucks |      |      |      |      |
|-----|------|------|------|------|------|--------------|------|------|------|------|
|     | 2011 | 2012 | 2013 | 2014 | 2015 | 2011         | 2012 | 2013 | 2014 | 2015 |
| $a$ | 38.2 | 40   | 40.8 | 41.2 | 41.7 | 30.9         | 32.7 | 34.1 | 34.1 | 34.3 |
| $b$ | 25.9 | 27.4 | 28.7 | 29.9 | 31.2 | 21.5         | 22.8 | 23.8 | 24.3 | 24.8 |
| $c$ | 45.9 | 45.8 | 45.7 | 45.6 | 45.5 | 51.9         | 52.0 | 52.0 | 52.1 | 52.1 |
| $d$ | 1.6  | 1.5  | 1.5  | 1.4  | 1.4  | 3.8          | 3.8  | 3.8  | 3.9  | 3.9  |

the constrained logistic function

$$\frac{1}{\kappa^T(\mathbf{y})} = \frac{1}{a} + \left( \frac{1}{b} - \frac{1}{a} \right) \left( \frac{e^{(\kappa^F(\mathbf{y})-c)/d}}{1 + e^{(\kappa^F(\mathbf{y})-c)/d}} \right) \quad (5.1)$$

where  $\kappa^F : \mathcal{Y} \rightarrow \mathbb{R}$  defines a vehicle's footprint from its characteristic vector and  $a, b, c$  are parameters chosen to obtain increasingly stringent standards over 2011-2015 (NHTSA, 2006, 2008a).

#### 5.2.4 Corporate Average Emissions (CAE) Standards.

In response to California's Assembly Bill 1493 (Pavley, 2002), California's Air Resources Board proposed to regulate the Greenhouse Gas emissions from vehicles (including CO<sub>2</sub>) via a Corporate Average Emissions (CAE) Standard (CARB, 2008). Let  $\kappa : \mathcal{Y} \rightarrow \mathbb{R}_+$  define a vehicle's CO<sub>2</sub> emissions intensity from its characteristic vector. A basic, single-class CAE standard has the penalty

$$c_f^R(\mathbf{q}, \mathbf{Y}) = c_f^R(\mathbf{q}_f, \mathbf{Y}_f) = (\mathbf{q}_f^\top \mathbf{1}) \max\{0, \mathbf{a}^A(\boldsymbol{\kappa}_f(\mathbf{Y}), \mathbf{q}_f) - \kappa_f^S\}$$

for a firm-specific emissions intensity standard level  $\kappa_f^S$  and where  $\bar{\kappa}_f^A(\mathbf{q}_f, \boldsymbol{\kappa}_f(\mathbf{Y}))$  is the arithmetic average of the values  $\boldsymbol{\kappa}_f(\mathbf{Y})$  weighted by  $\mathbf{q}_f$ . Thus the single-class CAE standard has heuristic expected regulatory costs

$$\hat{c}_f^R(\mathbf{p}) = r_f (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \max\{0, \bar{\kappa}_f^A(\mathbf{p}) - \kappa_f^S\} \quad \text{where} \quad \bar{\kappa}_f^A(\mathbf{p}) = \mathbf{a}^A(\boldsymbol{\kappa}_f, \mathbf{P}_f(\mathbf{p})).$$

and satisfies Assumption 5.1.2 with  $\boldsymbol{\varrho}_f(\mathbf{p}) = \max\{0, \bar{\kappa}_f^A(\mathbf{p}) - \kappa_f^S\} \mathbf{1}$ . Again  $\boldsymbol{\varrho}_f$  is not differentiable for any  $\mathbf{p}$  such that  $\hat{\kappa}_f^H(\mathbf{p}) = \kappa_f^S$ .

California's waiver for the legal authority to implement this standard was denied

by the federal government (Maynard, 2007), and thus this standard is not currently in place in the U.S. vehicle market.

### 5.2.5 Feebates.

Feebates are similar to the CAFE and CAE standard policies discussed above in that firms that violate a standard incur fees, but differ in that firms that satisfy the standard are rewarded with rebates (Greene et al., 2005). A feebate could be placed on any characteristic of interest; i.e. fuel economy, fuel consumption, or CO<sub>2</sub> emissions intensity. The form of regulatory costs depends on what characteristic is regulated.

#### Fixed Pivot Feebates

In placing a feebate on fuel economy, the government would assign a fixed “pivot” fuel economy  $\kappa^S$  such that firms are penalized for offering vehicles with fuel economies below this pivot and rewarded for having fuel economies above this limit. The heuristic expected regulatory costs for such a policy are

$$\hat{c}_f^R(\mathbf{p}) = \mathbf{P}_f(\mathbf{p})^\top \boldsymbol{\varrho}_f \quad \text{and} \quad \boldsymbol{\varrho}_f = \kappa^S \mathbf{1} - \boldsymbol{\kappa}_f.$$

More generally, the pivot could be made characteristic-dependent by specifying  $\kappa^S : \mathcal{Y} \rightarrow \mathbb{R}$ . The heuristic expected regulatory costs would then be generated by  $\boldsymbol{\varrho}$  with components  $\varrho_j = \kappa^S(\mathbf{y}_j) - \kappa_j$ .

In placing a feebate on fuel consumption (or emissions intensity), the government would again assign a fixed “pivot” fuel consumption  $\kappa_f^S$  such that firms are penalized for offering vehicles with fuel consumption above this pivot and rewarded for having fuel consumptions below this limit. The heuristic expected regulatory costs for such a policy are

$$\hat{c}_f^R(\mathbf{p}) = \mathbf{P}_f(\mathbf{p})^\top \boldsymbol{\varrho}_f \quad \text{and} \quad \boldsymbol{\varrho}_f = \boldsymbol{\kappa}_f - \kappa_f^S \mathbf{1}.$$

More generally, the pivot could be made characteristic-dependent by specifying  $\kappa^S : \mathcal{Y} \rightarrow \mathbb{R}$ . The heuristic expected regulatory costs would then be generated by  $\boldsymbol{\varrho}$  with components  $\varrho_j = \kappa_j - \kappa^S(\mathbf{y}_j)$ .

#### Revenue-Neutral Feebates

Feebates can be made “revenue neutral” instruments in the sense that they do not result in a net deficit or surplus of funds generated by the policy (ignoring administra-

tion costs) (Greene et al., 2005). This can be accomplished by setting  $r_f = r$  for all  $f \in \mathbb{N}(F)$  and using an appropriately chosen “flexible pivot”  $\kappa^S \equiv \kappa^S(\mathbf{p})$ .

**Proposition 5.2.1.**  $\sum_f \hat{c}_f^R(\mathbf{p}) \equiv 0$  if  $\kappa^S(\mathbf{p}) = \mathbf{a}^A(\boldsymbol{\kappa}, \mathbf{P}(\mathbf{p}))$ .

We remark that these flexible pivots give revenue-neutral policies not only for the heuristic expected regulatory costs, but for actual regulatory costs as well.

## 5.3 Stationarity

As in Chapters 3 and 4, the first-order stationary can be applied here when  $\boldsymbol{q}_f$  is differentiable. However, we have noted that some of the regulatory policies we consider are not continuously differentiable: the heuristic form of the CAFE, attribute-based CAFE, and CAE policies are not differentiable when the standards are exactly satisfied. In this section we provide straightforward necessary and sufficient conditions for profit-optimal prices under such standard policies below, based on the work of Clarke (1975) and Ioffe (1979). We also undertake some analysis based on these conditions, identifying basic incentives present under these regulatory policies and discussing “shadow taxes” on firms whose profit-optimal prices are implicitly constrained by a regulatory standard.

### 5.3.1 Smooth Stationarity

The following is a further adaptation of well-known necessary conditions for the local maximization of an unconstrained, continuously differentiable function (Munkres, 1991).

**Proposition 5.3.1.** *Let  $\mathcal{A} \subset (0, \varsigma_* \mathbf{1}) \subset \mathbb{R}^J$  be open. Suppose Assumptions 4.1.2 and 5.1.2 hold with  $\boldsymbol{q}_f$  is continuously differentiable on  $\mathcal{A}$ . Then  $\hat{\pi}_f^R$  is continuously differentiable on  $\mathcal{A}$ , and if  $\mathbf{p}_f$  maximizes  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$  on  $\mathcal{A}$  then  $(\nabla_f \hat{\pi}_f^R)(\mathbf{p}) = \mathbf{0}$  where*

$$(\nabla_f \hat{\pi}_f^R)(\mathbf{p}) = (D_f \mathbf{P}_f)(\mathbf{p})^\top (\mathbf{p}_f - \mathbf{c}_f - r_f \boldsymbol{q}_f(\mathbf{p})) + (\mathbf{I} - r_f (D_f \boldsymbol{q}_f)(\mathbf{p})^\top) \mathbf{P}_f(\mathbf{p}). \quad (5.2)$$

Eqn. (5.2) leads to a fixed-point characterization of profit-maximizing prices.

**Definition 5.3.1.** *We define the map  $\boldsymbol{\zeta}^R : \mathbb{R}^J \rightarrow \mathbb{R}^J$  through it's intra-firm components  $\boldsymbol{\zeta}_f^R : \mathbb{R}^J \rightarrow \mathbb{R}^{J_f}$ ,*

$$\boldsymbol{\zeta}_f^R(\mathbf{p}) = \boldsymbol{\zeta}_f(\mathbf{p}) - r_f \boldsymbol{\Lambda}_f(\mathbf{p})^{-1} (\boldsymbol{\Gamma}_f(\mathbf{p})^\top \boldsymbol{q}_f(\mathbf{p}) - (D_f \boldsymbol{q}_f)(\mathbf{p})^\top \mathbf{P}_f(\mathbf{p})) \quad (5.3)$$

wherever  $(D\boldsymbol{\varrho}_f)(\mathbf{p})$  exists. Otherwise,  $\boldsymbol{\zeta}_f^R$  can be defined arbitrarily.  $\boldsymbol{\zeta}$  is defined as in Chapter 4.

**Proposition 5.3.2.** *Suppose the hypotheses of Proposition 5.3.1.*

(i) *On  $\mathcal{A}$ , the regulated profit gradient can be written*

$$(\nabla_f \hat{\pi}_f^R)(\mathbf{p}) = \boldsymbol{\Lambda}_f(\mathbf{p})(\mathbf{p}_f - \mathbf{c}_f - r_f \boldsymbol{\varrho}_f(\mathbf{p}) - \boldsymbol{\zeta}_f^R(\mathbf{p})).$$

(ii) *If  $\mathbf{p}_f \in \mathcal{A}$  locally maximizes  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$ , then*

$$\mathbf{p}_f = \mathbf{c}_f + r_f \boldsymbol{\varrho}_f(\mathbf{p}) + \boldsymbol{\zeta}_f^R(\mathbf{p}). \quad (5.4)$$

## Incentives

These representations enable an analysis of the incentives engendered by different policies. In saying a firm has an “incentive” to raise (lower) the price of the  $j^{\text{th}}$  vehicle at prices  $\mathbf{p}$ , we simply mean that  $(D_j \hat{\pi}_f^R)(\mathbf{p}) > 0$  ( $< 0$ ).

**Proposition 5.3.3.** *The following hold:*

- (i) *Under the CO<sub>2</sub> tax, every firm has an incentive to increase the price of their most emissions-intensive vehicle, relative to unregulated profit-optimal prices.*
- (ii) *Suppose the regulation is a fixed-pivot feebate on fuel economy, and that for each firm  $f$  there are products  $k, l \in \mathcal{J}_f$  such that  $\kappa_k < \kappa^S < \kappa_l$ . Then every firm has an incentive to raise the price of its least efficient vehicle and lower the price of its most efficient vehicle relative to unregulated profit-optimal prices.*
- (iii) *Suppose the regulation is a fixed-pivot feebate on fuel consumption (or emissions intensity), and that for each firm  $f$  there are products  $k, l \in \mathcal{J}_f$  such that  $\kappa_k < \kappa^S < \kappa_l$ . Then every firm has an incentive to raise the price of its most fuel consumptive vehicle and lower the price of its least consumptive vehicle, relative to unregulated profit-optimal prices.*
- (iv) *Suppose that, under the single-class CAFE standards, firm  $f$ 's unregulated equilibrium prices violate the standard. Then from these prices the firm has an incentive to increase the price of its least fuel efficient vehicle.*
- (v) *Suppose that, under the single-class CAE standards, firm  $f$ 's unregulated profit-optimal prices violate the standard. Then from these prices the firm has an incentive to increase the price of its most emissions-intensive vehicle, and decrease the price of its most emissions-intensive vehicle.*

Claims (ii)-(v) are related to the “mix-shifting” incentives perceived by economists studying pricing in the automotive market under CAFE (Kleit, 1990; Goldberg, 1998;

Kleit, 2004). This form of these incentives are present for any Mixed Logit model for which the Leibniz Rule holds and under no assumptions on the firms or products offered by them. More is true than stated, in the sense that it will not only be the extremal vehicles that have prices shifting in the manners described. However it appears impossible to say in general exactly which vehicles, other than these extremal ones, will have prices affected in the stated ways without a numerical analysis.

### 5.3.2 Standards and Non-Smooth Stationarity

For continuously differentiable regulatory costs, the fixed-point expression of stationarity suffices. However these conditions must be modified to account for the (potential) non-differentiability of  $\hat{c}_f^R$  under standards like CAFE and CAE. We apply abstract necessary and sufficient conditions to this problem, reaching first and second order conditions satisfied by regulated profit-maximizing prices when the CAFE or CAE standards are satisfied exactly.

#### Gradient Conditions

The following proposition gives necessary and sufficient conditions applicable when  $\hat{q}_f(\mathbf{p}) = 0$ .

**Proposition 5.3.4.** *Suppose Assumption 5.1.3 holds and  $\mathbf{p}_f$  satisfies  $\hat{q}_f(\mathbf{p}_f, \mathbf{p}_{-f}) = 0$ .*

- (i) *If  $\mathbf{p}_f$  locally maximizes  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$  then there exists some  $v_f \in [0, r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})]$  such that  $(\nabla_f \hat{\pi}_f)(\mathbf{p}) = v_f(\nabla_f \hat{q}_f)(\mathbf{p})$ . Moreover, if  $v_f \in (0, r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}))$  then  $(D_f \nabla_f \hat{\pi}_f)(\mathbf{p}) - v_f(D_f \nabla_f \hat{q}_f)(\mathbf{p})$  is negative semi-definite on  $(\nabla_f \hat{q}_f)(\mathbf{p})^\perp$ .*
- (ii) *On the other hand if there exists some  $v_f \in (0, r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}))$  such that  $(\nabla_f \hat{\pi}_f)(\mathbf{p}) = v_f(\nabla_f \hat{q}_f)(\mathbf{p})$  and  $(D_f \nabla_f \hat{\pi}_f)(\mathbf{p}) - v_f(D_f \nabla_f \hat{q}_f)(\mathbf{p})$  is negative definite on  $(\nabla_f \hat{q}_f)(\mathbf{p})^\perp$ , then  $\mathbf{p}_f$  is a strict local maximizer of  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$ .*

The first-order necessary condition follows from the theory of generalized gradients developed by Clarke (1975), while the second-order necessary and the sufficient condition follow from those developed by Ioffe (1979). From a computational perspective, this formulation is beneficial because both (i) and (ii) are easy to test for a particular vector of prices  $\mathbf{p}_f$  satisfying  $\hat{q}_f(\mathbf{p}_f, \mathbf{p}_{-f}) = 0$ .

#### Shadow Taxes

Conceptual benefits of the characterization of standard-bound profit-optimal prices given in Proposition 5.3.4 also exist. Proposition 5.3.4 can be used to characterize the



“shadow” taxes structured by standard-based regulatory policies.

**Proposition 5.3.5.** *Suppose  $\mathbf{p}_f$  is stationary for  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$ , where  $\hat{\rho}_f(\mathbf{p}) = 0$ , and define  $\boldsymbol{\chi}_f(\mathbf{p})$  by  $(\nabla_f \hat{\rho}_f)(\mathbf{p}) = (D_f \mathbf{P}_f)(\mathbf{p})^\top \boldsymbol{\chi}_f(\mathbf{p})$ .<sup>4</sup> Then  $\mathbf{p}_f$  is stationary for the flat tax/subsidy regulated optimal pricing problem with flat taxes  $\mathbf{t}_f = v_f \boldsymbol{\chi}_f(\mathbf{p})$  where  $v_f \in [0, r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})]$  is the Lagrange multiplier from the non-smooth first-order conditions of the standard regulated problem.*

In this sense,  $v_f \boldsymbol{\chi}_f(\mathbf{p})$  are the “shadow,” or “effective,” taxes acting on firm  $f$  under the standard, when that standard is binding. By defining  $\boldsymbol{\chi}_f(\mathbf{p})$  for our standard policies, we can get a more specific handle on what drives these shadow taxes.

**Proposition 5.3.6.** *The following hold:*

(i) *Under the CAFE standards,*

$$\boldsymbol{\chi}_f(\mathbf{p}) = - \left( \frac{\bar{\kappa}_f^H(\mathbf{p})}{\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}} \right) \left( \mathbf{1} - \frac{\bar{\kappa}_f^H(\mathbf{p})}{\boldsymbol{\kappa}_f} \right).$$

(ii) *Under the attribute-based CAFE standards,*

$$\boldsymbol{\chi}_f(\mathbf{p}) = \left( \frac{1}{\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}} \right) \left[ \bar{\kappa}_f^S(\mathbf{p}) \left( \mathbf{1} - \frac{\bar{\kappa}_f^S(\mathbf{p})}{\boldsymbol{\kappa}_f^S(\mathbf{p})} \right) - \bar{\kappa}_f^H(\mathbf{p}) \left( \mathbf{1} - \frac{\bar{\kappa}_f^H(\mathbf{p})}{\boldsymbol{\kappa}_f} \right) \right]$$

(iii) *Under the CAE standards,*

$$\boldsymbol{\chi}_f(\mathbf{p}) = \left( \frac{1}{\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}} \right) (\boldsymbol{\kappa}_f - \bar{\kappa}_f^A(\mathbf{p}) \mathbf{1}).$$

The following intuitive claims are a consequence of Proposition 5.3.5 and Proposition 5.3.6.

**Corollary 5.3.7.** *The following hold:*

- (i) *Under the CAFE standard, any firm with standard-bound profit-optimal prices is effectively taxed for all vehicles with fuel economies below the standard and effectively subsidized for all vehicles exceeding the standard.*
- (ii) *Under the attribute-based CAFE standard, any firm with standard-bound profit-optimal prices is effectively taxed for all vehicles falling short of their target fuel economy and effectively subsidized for all vehicles exceeding their target fuel economy.*

---

<sup>4</sup> $\boldsymbol{\chi}_f$  is well-defined because  $(D_f \mathbf{P}_f)(\mathbf{p})$  is nonsingular.

(iii) Under the CAE standard, any firm with standard-bound profit-optimal prices is effectively taxed for all vehicles with emissions intensities above the standard and effectively subsidized for all vehicles with emissions intensities below the standard.

Because the standard-constrained optimization problems are independent of the penalty level  $r_f$ , we are also assured that shadow taxes exist whenever there are standard-bound profit maximizing prices.

**Corollary 5.3.8.** *Suppose  $\mathbf{p}_f$  is a strict local solution to the constrained optimization problem*

$$\begin{aligned} & \text{maximize} && \hat{\pi}_f(\mathbf{q}_f, \mathbf{p}_{-f}) \\ & \text{with respect to} && \mathbf{q}_f \in \mathbb{R}_+^{J_f} \\ & \text{subject to} && \hat{\varrho}_f(\mathbf{q}_f, \mathbf{p}_{-f}) = 0 \end{aligned} \tag{5.5}$$

with Lagrange multiplier  $v_f > 0$ . Then there exists a  $\bar{r}_f > 0$  such that  $\mathbf{p}_f$  locally maximizes  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$  for any penalty level  $r_f > \bar{r}_f$ .

The proof is trivial: any  $r_f > \bar{r}_f = v_f/(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})$  satisfies  $v_f < r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})$ .

### Fixed-Point Form

We can summarize the stationarity conditions for standards in fixed-point form using a revised definition of  $\zeta$ .

**Definition 5.3.2.** *Let*

$$\zeta_f^{R,*}(\mathbf{p}, v) = \mathbf{\Lambda}_f(\mathbf{p})^{-1}(\mathbf{\Gamma}_f(\mathbf{p})^\top \mathbf{m}_f^R(\mathbf{p}, v) - \mathbf{P}_f(\mathbf{p}))$$

where  $\mathbf{m}_f^R(\mathbf{p}, v) = \mathbf{p}_f - \mathbf{c}_f - r_f \max\{0, \hat{\varrho}_f(\mathbf{p})\} \mathbf{1} - v \mathbf{\chi}_f(\mathbf{p})$ .

**Proposition 5.3.9.** *Suppose  $\mathbf{p}_f$  locally maximizes  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$ . Then*

$$\mathbf{p}_f = \mathbf{c}_f + r_f \max\{0, \hat{\varrho}_f(\mathbf{p})\} \mathbf{1} + v_f \mathbf{\chi}_f(\mathbf{p}) + \zeta_f^{R,*}(\mathbf{p}, v_f) \tag{5.6}$$

where

$$v_f \begin{cases} = 0 & \text{if } \hat{\varrho}_f(\mathbf{p}) < 0 \\ \in [0, r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})] & \text{if } \hat{\varrho}_f(\mathbf{p}) = 0 \\ = r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) & \text{if } \hat{\varrho}_f(\mathbf{p}) > 0 \end{cases} .$$

### 5.3.3 Simultaneous Stationarity

Without regulation, or indeed with any regulatory policies having continuously differentiable  $\boldsymbol{q}_f$ , local equilibrium prices  $\mathbf{p}$  simultaneously satisfy Eqn. (5.4) for all  $f \in \mathbb{N}(F)$ . For non-differentiable standard policies satisfying Assumption 5.1.3, local equilibrium prices  $\mathbf{p}$  simultaneously satisfy Eqn. (5.6).

## 5.4 A Hybrid Fixed-Point Iteration

The existence of the fixed-point characterization of simultaneous stationarity and the successful implementation of a fixed-point iteration for unregulated problems suggests using fixed-point iteration to solve for regulated equilibrium prices. This is a straightforward application of Eqn. (5.3), and thus we do not discuss this here. Preliminary computations demonstrate that this approach works quite well; see Fig. 5.2.

However, the fixed-point iteration, or indeed any Newton-type method, cannot appropriately solve the non-smooth regulated equilibrium pricing problem under standards without fixing a collection of firms assumed to be bound by the standard. In this section we suggest a hybrid fixed-point iteration by adding direct solution of the non-smooth first-order conditions for those firms that are suspected of having standard-bound profit-maximizing prices. Moreover, we derive these suspicions from fixed-point iteration. This hybrid fixed-point iteration is thus akin to an active set strategy, in which algorithm process informs the designation of constraints as active.

Specifically, the following result leads us to expect that “naive” fixed-point iteration will tend to cycle near standard-bound profit-optimal prices.

**Proposition 5.4.1.** *Let  $\mathbf{p}_f$  locally maximize  $\hat{\pi}_f^R(\cdot, \mathbf{p}_{-f})$  where  $\hat{\boldsymbol{q}}_f(\mathbf{p}_f, \mathbf{p}_{-f})$ . Then  $\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) - \mathbf{p}_f$  and*

$$\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) + r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\boldsymbol{\Lambda}_f(\mathbf{p})^{-1}(\nabla_f \hat{\boldsymbol{q}}_f)(\mathbf{p}) - \mathbf{p}_f$$

*are colinear and point in opposite directions. That is, there is a scalar  $\alpha_f^R > 0$  such that*

$$\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) + r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\boldsymbol{\Lambda}_f(\mathbf{p})^{-1}(\nabla_f \hat{\boldsymbol{q}}_f)(\mathbf{p}) - \mathbf{p}_f = -\alpha_f^R (\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) - \mathbf{p}_f).$$

Moreover  $(\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) - \mathbf{p}_f)^\top (\nabla_f \hat{\varrho}_f)(\mathbf{p}) > 0$  and, by the first result,

$$\left( \mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) + r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\Lambda}_f(\mathbf{p})^{-1} (\nabla_f \hat{\varrho}_f)(\mathbf{p}) - \mathbf{p}_f \right)^\top (\nabla_f \hat{\varrho}_f)(\mathbf{p}) < 0.$$

A consequence of this result is that for prices  $\mathbf{p}'_f$  near  $\mathbf{p}_f$  but with  $\hat{\varrho}_f(\mathbf{p}'_f, \mathbf{p}_{-f}) < 0$ , the fixed-point step points towards violation of the standard, while for prices  $\mathbf{p}'_f$  near  $\mathbf{p}_f$  but with  $\hat{\varrho}_f(\mathbf{p}'_f, \mathbf{p}_{-f}) > 0$ , the fixed-point step points towards satisfaction of the standard. These fixed point steps do not vanish as exact satisfaction of the standard is approached.

We can take this observation as the basis of a hybrid algorithm for solving unconstrained price equilibrium problems under standards. For any prices that either violate or satisfy a firm's standard function, we simply take the fixed-point step prescribed by Eqn. (5.3). Particularly, if the current prices are such that firm  $f$  satisfies the standard, the unregulated fixed-point step is taken. If the corresponding sequence of iterates displays some firm that tends to set prices that generate a cycle between satisfying and violating the standard, then we take this to imply that the firm actually has profit-maximizing prices that exactly satisfy the standard. Such prices cannot be reached by fixed-point steps (at least without some form of damping), and thus in future iterations we solve the first-order conditions instead of taking fixed-point steps.

Suppose then that the firms indexed by  $\mathcal{F} \subset \mathbb{N}(F)$  are suspected of having standard-bound equilibrium prices. Let  $\mathbf{p}_{\mathcal{F}}$  be that sub-vector of  $\mathbf{p}$  containing only the prices for firms in  $\mathcal{F}$ , and  $\mathbf{p}_{-\mathcal{F}}$  be that sub-vector of  $\mathbf{p}$  containing only the prices for firms in  $\mathbb{N}(F) \setminus \mathcal{F}$ . We solve for a *simultaneous* solution  $(\mathbf{p}_{\mathcal{F}}, \mathbf{v}_{\mathcal{F}}) \in \prod_{f \in \mathcal{F}} \mathbb{R}^{J_f} \times \mathbb{R}^{|\mathcal{F}|}$  to the first-order conditions

$$\begin{aligned} (\nabla_f \hat{\pi}_f)(\mathbf{p}_{\mathcal{F}}, \mathbf{p}_{-\mathcal{F}}) &= v_f (\nabla_f \hat{\varrho}_f)(\mathbf{p}_{\mathcal{F}}, \mathbf{p}_{-\mathcal{F}}) & \text{for all } f \in \mathcal{F}. \\ \hat{\varrho}_f(\mathbf{p}_{\mathcal{F}}, \mathbf{p}_{-\mathcal{F}}) &= 0 \end{aligned} \tag{5.7}$$

Newton's method can be used to solve this system on equations. The bounds  $0 \leq v_f \leq r_f(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})$  must also be checked to validate that indeed a simultaneously stationary point is found.

Given the success of the fixed-point approach  $\boldsymbol{\zeta}$ -NM for unregulated price equilibrium computations, we propose a hybrid fixed-point algorithm outlined in Fig. 5.1 using the fixed-point form of the conditions (5.7). In fact, preliminary computational experiments have demonstrated that solving the conditions (5.7) in their given, combined gradient form can be very difficult. To use the fixed-point form we define

**Figure 5.1** A hybrid fixed-point approach for standard-regulated Bertrand competition. The following steps update  $\mathbf{p}$ , an estimate of regulated equilibrium prices, and  $\mathcal{F}$ , an estimate of the firms standard-bound in equilibrium.

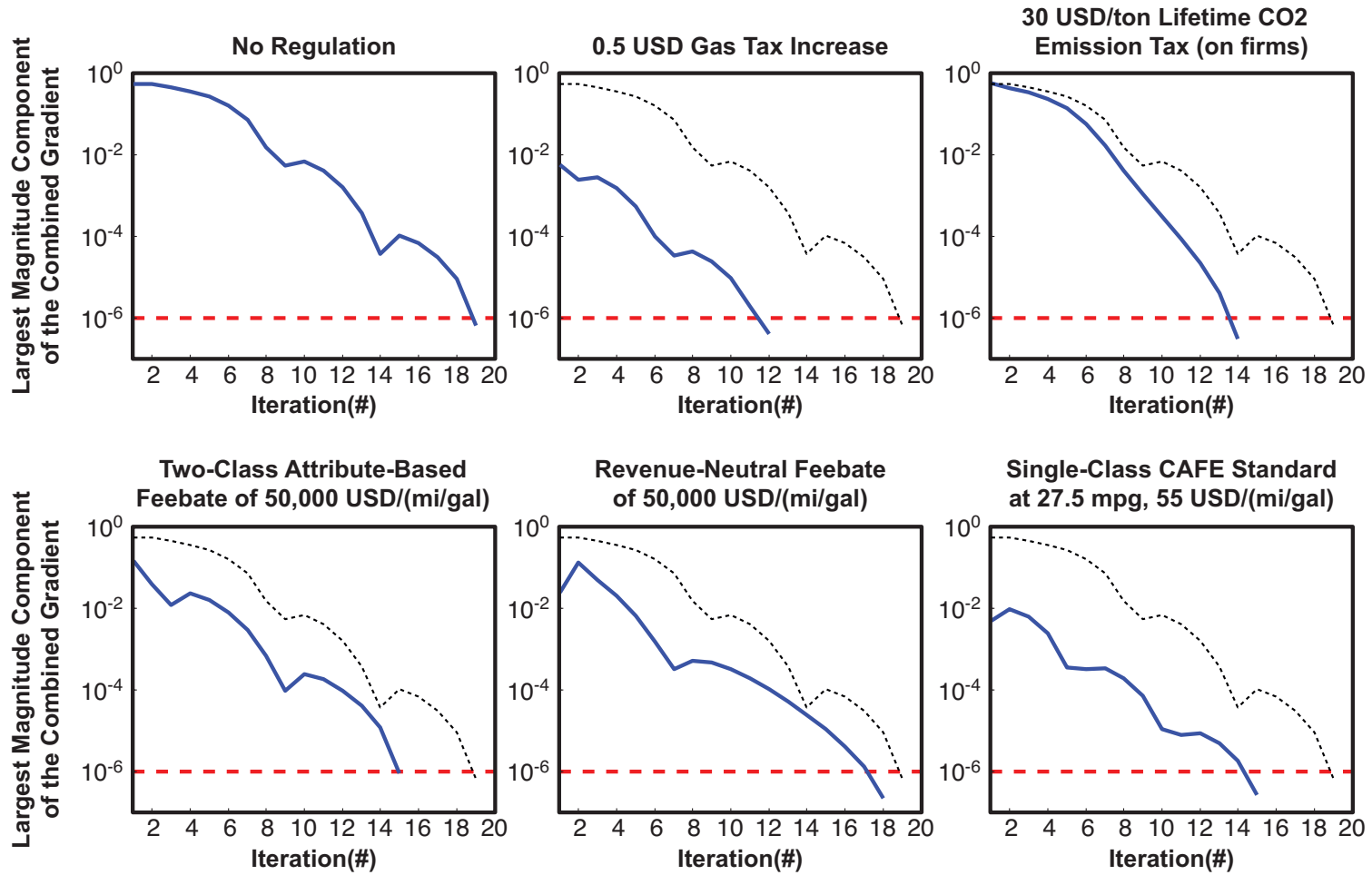
- ▷ compute  $\boldsymbol{\zeta}$ ,  $\hat{\boldsymbol{\rho}} = \{\hat{\rho}_f\}_{f=1}^F$ ,  $\boldsymbol{\rho}$ ,  $\boldsymbol{\zeta}^R$
- ▷ terminate if the first-order conditions are satisfied
- ▷ append firms  $f \notin \mathcal{F}$  suspected of being standard-bound to  $\mathcal{F}$
- ▷ for  $f \notin \mathcal{F}$ , compute  $\mathbf{q}_f \leftarrow \mathbf{c}_f + r_f \boldsymbol{\rho}_f + \boldsymbol{\zeta}_f^R$
- ▷ if  $\mathcal{F} \neq \{\emptyset\}$ , compute  $(\mathbf{q}_{\mathcal{F}}, \mathbf{v}_{\mathcal{F}})$ , a Newton step for  $\boldsymbol{\varphi}_{\mathcal{F}}$
- ▷ update  $\mathbf{p} \leftarrow \mathbf{q}$

$\boldsymbol{\varphi}_{\mathcal{F}}^p : \prod_{f \in \mathcal{F}} \mathbb{R}^{J_f} \times \mathbb{R}^{|\mathcal{F}|} \rightarrow \prod_{f \in \mathcal{F}} \mathbb{R}^{J_f}$  and  $\boldsymbol{\varphi}_{\mathcal{F}}^\sigma : \prod_{f \in \mathcal{F}} \mathbb{R}^{J_f} \times \mathbb{R}^{|\mathcal{F}|} \rightarrow \mathbb{R}^{|\mathcal{F}|}$  by

$$\begin{aligned} \varphi_k^p(\mathbf{p}_{\mathcal{F}}, \boldsymbol{\sigma}_{\mathcal{F}}) &= (\boldsymbol{\varphi}_{\mathcal{F}}^p(\mathbf{p}_{\mathcal{F}}, \boldsymbol{\lambda}_{\mathcal{F}}))_k \\ &= p_k - c_k - v_{f(k)} \chi_k(\mathbf{p}_{\mathcal{F}}, \mathbf{p}_{-\mathcal{F}}) - \zeta_k^{R,*}(\mathbf{p}_{\mathcal{F}}, \mathbf{p}_{-\mathcal{F}}, v_{f(k)}) \text{ for all } k \in \mathcal{J}_{\mathcal{F}} \\ \varphi_f^\sigma(\mathbf{p}_{\mathcal{F}}, \boldsymbol{\sigma}_{\mathcal{F}}) &= (\boldsymbol{\varphi}_{\mathcal{F}}^\sigma(\mathbf{p}_{\mathcal{F}}, \boldsymbol{\sigma}_{\mathcal{F}}))_f \\ &= \hat{\rho}_f(\mathbf{p}_{\mathcal{F}}, \mathbf{p}_{-\mathcal{F}}) \text{ for all } f \in \mathcal{F}. \end{aligned}$$

We then define  $\boldsymbol{\varphi}_{\mathcal{F}} : \prod_{f \in \mathcal{F}} \mathbb{R}^{J_f} \times \mathbb{R}^{|\mathcal{F}|} \rightarrow \prod_{f \in \mathcal{F}} \mathbb{R}^{J_f} \times \mathbb{R}^{|\mathcal{F}|}$  by  $\boldsymbol{\varphi}_{\mathcal{F}} = (\boldsymbol{\varphi}_{\mathcal{F}}^p, \boldsymbol{\varphi}_{\mathcal{F}}^\sigma)$ . A pair  $(\mathbf{p}_{\mathcal{F}}, \mathbf{v}_{\mathcal{F}}) \in \prod_{f \in \mathcal{F}} \mathbb{R}^{J_f} \times \mathbb{R}^{|\mathcal{F}|}$  simultaneously solves the first-order conditions (5.7) if, and only if, it is a zero of  $\boldsymbol{\varphi}_{\mathcal{F}}$ .

**Figure 5.2** Convergence curves for several regulated fixed-point iterations under the [Berry et al. \(1995\)](#) model. Unregulated fixed-point iteration started at unit costs, and regulated fixed-point iterations started at unregulated equilibrium prices. Under CAFE, no firms are bound by the standard.



## 5.5 Conclusions

In this chapter we have extended our fixed-point approach from Chapters 3 and 4 to the setting where firms face regulatory costs. We have devised a general framework under which regulated Bertrand competition can be analyzed for any market, but focused on the automotive industry for specific policies. While a number of recent analyses consider the impact of regulations on Bertrand competition, many of these study only policies that influence consumer utility rather than firm profit functions directly. Several analyses of the CAFE standards that do include regulation's impact on firm's profit functions study a single policy and fix which firms behave as if bound by the standards. In order to relax these assumptions, we have adopted necessary and sufficient conditions from nonsmooth optimization and proposed a hybrid fixed-point iteration for the computation of regulated equilibrium prices. While preliminary computations using the extended fixed-point approach have been successful, future work needs to present a detailed study of the numerical properties of these algorithms similar to that presented in Chapter 4. Our theory also needs to be extended to multi-class standards.

# Chapter 6

## Conclusions and Future Work

This dissertation has presented several contributions to the existing theory of equilibrium pricing. We have described a cohesive framework for Bertrand competition under an arbitrary RUM. The development of this framework has identified fundamental assumptions that may currently be overlooked in prominent econometric applications of Bertrand competition. We have also undertaken a detailed analysis of equilibrium prices under the Logit RUM, resulting in a new proof of the existence of equilibrium that does not rely on restrictive assumptions on the number and type of products or firms. Our analysis has also uncovered one counterintuitive property of profit-optimal prices under many Logit models, that markups are constant or decrease with value. We have developed and experimented with a fixed-point approach for computing equilibrium prices in large-scale and complex differentiated product markets with Mixed Logit RUMs. Despite the prominence of such models, focused study of the computation of equilibrium prices has not yet been undertaken. We have clearly demonstrated two benefits to the fixed-point approach to equilibrium prices:

- (i) “Analytically preconditioning” the equilibrium problem by using  $\eta$ -NM or  $\zeta$ -NM instead of CG-NM yields a significant reduction in the time required to compute equilibrium prices as well as an improvement in reliability;
- (ii)  $\zeta$ -FPI can be a very reliable and efficient method of computing equilibrium prices.

Due to the success of the general fixed-point approach for computing equilibrium prices without regulation, we have also extended the fixed-point approach to models where firms face regulatory policies. All of these results contribute to the theory of Bertrand competition in differentiated product markets applied by economists, marketers, and engineers in ways that have been outlined in Chapter 1.



In this concluding chapter, we critically reflect on these contributions and pose directions for future work.

## 6.1 Criticisms of this Research

This research focuses entirely on Bertrand competition models of differentiated product markets where demand is characterized using a Logit or Mixed Logit RUM. While such models have played a prominent role in econometric studies of the automotive market for over a decade, they have not been completely validated as appropriate models of this market. An argument might therefore be made that this research focuses on a mathematical framework that is not useful for policy purposes, simply because it does not truly represent real markets.

We agree that future work needs to more rigorously validate particular Bertrand competition models as representative of real markets, and contend that the research developed in this dissertation already helps achieve that goal. For example, in Chapter 3, we identified *specific* ways in which Bertrand competition may be an inappropriate model of real markets. The decreasing-in-cost (or -quality) equilibrium markups under any reasonable Logit model is a direct, qualitative conclusion that can be succinctly verified or rejected with real market data. This particular example may seem unnecessary given the anecdotal prominence of markups that increase with quality. One item of future work discussed below is the extension of this analysis to Mixed Logit models, where these ideas may find more empirical relevance. However, it is likely that a direct analysis of Mixed Logit models will be limited to specific, simple cases. This is where our second contribution may play a role. The development of fast and reliable methods of computing equilibrium prices enables the exploration of complex Bertrand competition models that cannot be analyzed directly. With these methods, computational experiments can be undertaken that will allow practitioners to explore the pricing implications of alternative models, and whether these are plausible or counterintuitive. Such examinations have not yet been undertaken.

A more technical criticism that has been made of this research is that the numerical studies reported in Chapter 4 are not based on widely-available, professionally-developed software packages for solving nonlinear equations. We have used our own programs developed in `matlab` in order to run comparisons of various approaches on a “level playing field.” That being said, the GMRES-Hookstep (Viswanath, 2007, 2008), one elegant, efficient, and robust globally-convergent variant of Newton’s method for large-scale price equilibrium problems, is not currently available in any profes-

sional software package. As part of this work, we have developed a generic `matlab` implementation of this method.<sup>1</sup> We have also compared the computational times required of our methods with those of `matlab`'s professional nonlinear equations solver `fpsolve`, finding our GMRES-Hookstep method to compute solutions in comparable amounts of time. Use of both directional finite differences and analytical Jacobians, as well as the use of three distinct formulations, serves as an extensive “validity check.” Implementations of the fixed-point iteration and the GMRES-Hookstep methods in FORTRAN and/or C should be somewhat straightforward to develop, at which point formal comparisons between professional solvers like MINPACK (MINPACK, 2008), KNITRO (Byrd et al., 2006; KNITRO, 2008), KINSOL (Brown and Saad, 1990; KINSOL, 2008), or NITSOL (Pernice and Walker, 1998; Walker, 2005) our new methods can be undertaken. However, there is no reason to expect that the conclusions of our study would change.

Finally, our conclusion that existing results may be sensitive to the sample set may be critiqued for examining results under the  $\zeta$ -FPI only, rather than with all the numerical approaches considered. Certainly the same experiments should be undertaken with all methods to truly gauge the variation resulting from sample set selection. However, in all our comparison trials the prices computed by all methods were highly consistent, far more so than the results of equilibrium price computations under different sample sets. This suggests that where one method ( $\zeta$ -FPI) is sensitive, the others (CG-NM,  $\eta$ -NM, and  $\zeta$ -NM) will be as well.

## 6.2 Future Work

We divide elements of future work into two categories: those concerning price equilibrium and those considering single- and two-stage design and pricing games.

### 6.2.1 Price Equilibrium

#### Uniqueness of Equilibrium Prices

One key theoretical aspect of equilibrium pricing that we have not addressed is uniqueness. However, this is a very important property to prove in dealing with empirical Bertrand competition models of differentiated product markets. Multi-product firm

---

<sup>1</sup>Our Householder GMRES implementation matches `matlab`'s implementation step-by-step, and we have made use of the primary aspects of existing trust region methods as expounded by Dennis and Schnabel (1996).

equilibrium prices are known to be unique only under a linear in price utility Logit model (Sandor, 2001). We can currently perceive three potential avenues towards a uniqueness proof that extends to nonlinear utility functions. First, the most powerful result would be to prove contractivity of the map  $\mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$ . This would not only prove equilibrium uniqueness but also establish the fixed-point iteration as a globally convergent computational method. We address this issue further below. Second, Kellogg (1975) has proven the following result:

**Theorem 6.2.1.** (*Kellogg (1975)*) *Let  $\mathcal{X}$  be a convex compact subset of a finite dimensional Banach space, and let  $\varphi : \mathcal{X} \rightarrow \mathcal{X}$  be a continuously differentiable map such that (i) 1 is not an eigenvalue of  $(D\varphi)(x)$  for any  $x \in \mathcal{X}$  and (ii) no  $x \in \partial\mathcal{X}$ , the boundary of  $\mathcal{X}$ , is a fixed-point of  $\varphi$ . Then  $\varphi$  has a unique fixed-point in  $\mathcal{X}$ .*

In the context of Logit models this theorem could be applied to  $\mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$ ,  $\mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$ , or to the best response map (whose Jacobian can be derived from the Chain Rule). Note that unlike contractivity, Kellogg's theorem says nothing about the convergence of fixed-point iteration. Third, it might be possible to apply the Poincare-Hopf theorem yet again to prove uniqueness of equilibrium prices if the index of any zero of  $\mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p})$  or  $\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p})$  can be computed. Particularly, if the index of any such zero can be shown to be one, then the Poincare-Hopf theorem will again prove uniqueness of equilibria. This feels like a strong property to prove, but it may nonetheless be possible. Both the second and third approach must be based on a more developed understanding of  $(D\boldsymbol{\eta})(\mathbf{p})$  or  $(D\boldsymbol{\zeta})(\mathbf{p})$ .

## Convergence of the Fixed-Point Iteration

The most striking and unexplained technical outcome of this research is the fast and virtually ubiquitous convergence of the fixed-point iteration  $\mathbf{p} \leftarrow \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  for the specific models we explored. In Lemma 4.1.5, Claim (i) we stated that  $\|\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top\|_\infty < 1$ . Again, the proof of this claim is based on the simple observation that the  $k^{\text{th}}$  row sum of  $\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top$  equals

$$\int \left( \sum_{j \in \mathcal{J}_f(k)} P_j^L(\boldsymbol{\theta}, \mathbf{p}) \right) d\mu_{k,\mathbf{p}}(\boldsymbol{\theta}) < 1$$

where

$$d\mu_{k,\mathbf{p}}(\boldsymbol{\theta}) = \mu_k(\boldsymbol{\theta}, \mathbf{p})d\mu(\boldsymbol{\theta}) \quad \mu_k(\boldsymbol{\theta}, \mathbf{p}) = \frac{|(Dw_k)(\boldsymbol{\theta})| P_k^L(\boldsymbol{\theta}, \mathbf{p})}{\int |(Dw_k)(\boldsymbol{\theta}')| P_k^L(\boldsymbol{\theta}', \mathbf{p})d\mu(\boldsymbol{\theta}')}$$

Based on this equation, we also conjectured that  $\|\mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top\|_\infty$  should be reasonably bounded away from one. Combined with the appearance of  $\mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top(\mathbf{p} - \mathbf{c})$  in  $\boldsymbol{\zeta}(\mathbf{p})$ , this fact makes it seem somewhat plausible that  $\boldsymbol{\zeta}$  will have some contractivity. Investigating this further, both near equilibrium and globally, may yield strong theoretical support for use of the fixed-point iteration to supplement the numerical evidence given in this dissertation.

## Further Development of Numerical Methods

One direct item for future work is the development of codes for the  $\boldsymbol{\zeta}$ -FPI and a generic version of the GMRES-Hookstep in  $\mathbb{C}$  taking full advantage of the Basic Linear Algebra Subroutines (“BLAS”) (BLAS, 2008). Once this is completed, a comprehensive comparison of the different approaches for computing equilibrium prices can be undertaken with professionally developed codes like MINPACK, KNITRO, KINSOL, and NITSOL. We do not, however, anticipate that such study will change the qualitative outcomes of our existing comparisons:  $\boldsymbol{\eta}$ -NM and  $\boldsymbol{\zeta}$ -NM are superior to CG-NM and  $\boldsymbol{\zeta}$ -FPI can be quite efficient and reliable.

There are three areas where numerical methods proposed in this dissertation might be further improved. First, the success of  $\boldsymbol{\zeta}$ -FPI suggests “series acceleration” techniques (Weniger, 1989) as a means to further speed up computations. These methods can achieve the quadratic convergence rate of Newton’s method using compounded fixed-point steps. A related approach, which also connects to quasi-Newton methods for fixed-point problems, is Steffensen’s iteration (Johnson and Scholz, 1968; Ortega and Rheinboldt, 1970). Second, most of the computational burden of any method we have employed is derived from the integral approximations involved in computing utilities, choice probabilities, and choice probability derivatives. Alternative integral approximation methods such as importance or quasi-random sampling (Train, 2003; Train and Winston, 2007) could reduce the number of samples required to accurately compute equilibrium prices. Alternatively, parallelism could be exploited in the computation of equilibrium prices in both the computation of utilities, choice probabilities, and choice probability derivatives and the execution of steps based on these quantities.

## Analysis of Mixed Logit Models

Analysis of equilibrium prices under Mixed Logit models is warranted as one path to validating Bertrand competition as a practical model of real markets. The first and most important question, one that cannot be addressed through computations, is the

existence of equilibrium. We expect that the Poincare-Hopf Theorem will again imply the existence of fixed-points given the right assumption on the asymptotic rate of decrease of the utility function with respect to price. However the claim that stationary prices are profit maximizing, the central piece of the argument for Logit models, is obscured under a generic Mixed Logit model. It is possible that classes of models can be found for which this claim can be proven, or that a more involved degree-theoretic argument can be made to prove the existence of equilibrium. A particularly useful result here would be to find an example model where equilibrium does *not* exist. Such a finding would begin to structure the assumptions on  $(u, \vartheta)$  necessary for equilibrium to exist.

The new fixed-point expression derived in this dissertation also isolates the ability to generalize equilibrium markups by “mixing” Logit models. Specifically,  $\mathbf{p} = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p})$  is componentwise equivalent to

$$p_k = c_k + \int \zeta_k^L(\boldsymbol{\theta}, \mathbf{p}) d\mu_{k,\mathbf{p}}(\boldsymbol{\theta}) \quad \text{where} \quad \zeta_k^L(\boldsymbol{\theta}, \mathbf{p}) = \hat{\pi}_{f(k)}^L(\boldsymbol{\theta}, \mathbf{p}) - (Dw_k)(\boldsymbol{\theta}, p_k)^{-1}.$$

It is the product-and-price dependent measures  $d\mu_{k,\mathbf{p}}(\boldsymbol{\theta})$  that relieve the reductive properties of Logit equilibrium prices discussed in Chapter 3. Particularly, if for some  $k, l \in \mathcal{J}_f$   $d\mu_{k,\mathbf{p}} \equiv d\mu_{l,\mathbf{p}}$  for all  $\mathbf{p}$ , then the same properties of the two prices  $p_k, p_l$  and markups  $p_k - c_k, p_l - c_l$  proved for Logit pass over to the Mixed Logit setting. In other words, the the measures  $d\mu_{k,\mathbf{p}}$  isolate the ability of a Mixed Logit specification to represent more realistic markup patterns than those that hold under Logit. Further study of these measures will thus play a fundamental role in study of the properties of equilibrium prices under Mixed Logit models. First of all, whether or not  $d\mu_{k,\mathbf{p}} \equiv d\mu_{l,\mathbf{p}}$  can *ever* be true for a “nontrivial” Mixed Logit specification<sup>2</sup> remains an open question.

## Generalized Extreme Value Models

Generalized Extreme Value models (Train, 2003) are another prominent class of RUMs used to model the automotive industry. Loosely speaking, these models generalize Logit models by introducing across-product correlations in random utilities. The

---

<sup>2</sup>A “trivial” Mixed Logit specification would be one having  $w$  and  $v$  *both* independent of  $\boldsymbol{\theta}$ . By exclusion, a “nontrivial” Mixed Logit specification has one of  $w$  and  $v$  dependent on  $\boldsymbol{\theta}$ . If only  $\vartheta$  depends on  $\boldsymbol{\theta}$ , then  $d\mu_{k,\mathbf{p}} = d\mu_{l,\mathbf{p}}$  for all  $\mathbf{p}$  and all  $k, l$ . Thus a nontrivial Mixed Logit model must at least have price or characteristic components of utility that depend on the demographics. However,  $d\mu_{k,\mathbf{p}} \neq d\mu_{l,\mathbf{p}}$  for some  $\mathbf{p}$  even for simple cases such as  $w(\boldsymbol{\theta}, \mathbf{p}) = -\alpha p$  with  $v_k(\boldsymbol{\theta})$  anything not constant in  $\boldsymbol{\theta}$ . This raises some skepticism as to whether  $d\mu_{k,\mathbf{p}} = d\mu_{l,\mathbf{p}}$  can ever hold for nontrivial Mixed Logit models.

fixed-point approach described here extends to these models in the form

$$(\mathbf{I} + \mathbf{G}(\mathbf{p}))(\mathbf{p} - \mathbf{c}) = \boldsymbol{\zeta}(\mathbf{p})$$

for a specific matrix  $\mathbf{G}(\mathbf{p})$  related to these across-product correlations in utility. One economist has already expressed an interest in developing the analysis and computational experience with this model class.

## Multi-Class Regulatory Policies

Including multi-class policies is arguably the important extension of the methods discussed in Chapter 5. Without multi-class policies, the actual regulations in place in the U.S. today cannot be analyzed. While this extension may make numerical solution of the first-order conditions more challenging to program and implement, this extension may not involve a more challenging theory. The adopted results from non-smooth analysis and the primary numerical methods certainly apply to models with multi-class standards.

### 6.2.2 Single- and Multi-Stage Design Games

While this dissertation has dealt solely with price equilibrium, equilibrium designs and prices will play an important role in game-theoretic models of differentiated product markets for policy analysis.

#### Single-Stage Design Games

Simple single-stage design games can also be analyzed using the stationarity conditions. For the simplest models of the type introduced in Chapter 1: single-stage games of unconstrained characteristics choice ( $\mathbf{x} \equiv \mathbf{y}$ ,  $\mathcal{Y} \sim \mathbb{R}^K$ ) differentiable unit costs ( $c_f^U : \mathcal{Y} \rightarrow (0, \infty)$ ) and without fixed costs ( $c_f^F(\mathbf{Y}_f) \equiv c_f^F \geq 0$ ), a very conceptually appealing representation of equilibrium designs can be shown to hold: for each product, the gradient of unit costs with respect to characteristics is equal to the local willingness to pay for changes to these characteristics; i.e.

$$(\nabla^y c_f^U)(\mathbf{y}_k) = \frac{(\nabla^y u)(\mathbf{y}_k, p_k)}{|(D^p u)(\mathbf{y}_k, p_k)|} = \frac{(\nabla^y w)(\mathbf{y}_k, p_k) + (\nabla^y v)(\mathbf{y}_k)}{|(D^p w)(\mathbf{y}_k, p_k)|}.$$

The conceptual appeal can be described as follows: If this equation were to *fail*, then there would exist a characteristics change that consumers were willing to pay *more* for

than the incremental unit cost of the change to the firm. In particular, this expression shows that linear in price utility Logit models ( $w(\mathbf{y}, p) = -\alpha(\mathbf{y})p$ ) may generate optimal characteristics choices that are independent of competitor actions.<sup>3</sup> We say “may” because it is easy to use this equation to also derive specifications for  $c_f^U$  and  $v$  for which no profit-optimal characteristics choices exist. Furthermore, it is fairly easy to see that this criteria holds not just for Logit models, but for any GEV model with across-product correlations in error that are independent of product characteristics. In other words, while adding across-product correlations in the additive errors can result in models with more realistic choice predictions, this does not necessarily result in well-posed or realistic product design behavior.

Taking the engineering perspective, we could allow these potentially ill-posed models to be used recognizing that technical constraints subsequently determine the optimal solutions. While this may be an appropriate theoretical resolution, the subsequent question for economic theory is whether it will allow engineering considerations to make its models well-posed or if it would rather change the modeling paradigm to have models that are inherently well-posed.

## Two-Stage Design Games

Two-stage design and pricing games are an alternative modeling construct preferred by theoretical economists that may “naturally” resolve this potential ill-posedness. Future research must undertake a basic analysis of the two-stage model. A first step would be to determine if and when the two-stage model is in fact *different* from the single-stage model. The central obstacle to such an analysis is a characterization of the derivatives of equilibrium prices with respect to changes in design variables. Abstract formulae for these derivatives are easily derived from the Chain Rule by invoking the Implicit Function Theorem (Munkres, 1991), but understanding how they impact the gradient of profits as a function of design decisions alone, given prices in equilibrium, is a more challenging task for analysis.

Such abstract formulae can, however, be implemented in computations. These computations can immediately capture quantities of interest like shadow profits corresponding to changes in unit costs, product characteristics, or regulatory policy parameters. The use of these formulae in two-level optimization or two-stage games of regulated design and pricing is also feasible, and should occupy a central place in further work.

---

<sup>3</sup>Simply note that both sides of the given equation are functions of  $\mathbf{y}_k$  only.

When considering numerical approaches for two-stage games of regulated design and pricing, we must recall a guiding principle for this dissertation: *finding the right formulation for a problem makes all the difference in its solution*. While fixed-point expressions will not exist for design decisions as they have for prices,<sup>4</sup> there may still exist reformulations of the stationarity conditions for design problems that will be important for efficient and reliable computations. Our use of the hybrid fixed-point form of the non-smooth stationarity conditions for regulated equilibrium prices is an example of this.

### 6.2.3 A Roadmap for Policy Analysis

Quantitative policy analysis should be *informative, scientific, practical, complete, and effectively communicated*. Informative analyses address questions of relevance to the policy debate. Scientific analyses are based on principles verified using logic and empirical observations; in other words, based on sound science. Practical analyses can be reasonably undertaken and replicated. Complete analyses address all features of the system being regulated that have a reasonable potential to be influential in determining policy outcomes. Finally, an analysis is effectively communicated when influential policy makers understand the implications of the analysis for policy decisions.

A reasonable degree of completeness requires considering all of the disparate aspects considered in the various analyses outlined in Chapter 1 including: heterogeneity in consumer preferences, both for vehicle purchasing and driving; heterogeneous, self-interested firms; vehicle purchasing decisions including new and used cars, as well as scrappage decisions; vehicle-specific use decisions included alongside vehicle purchasing decisions. Specifically, game-theoretic models of regulated design and pricing are intended to make policy analysis more complete by depicting the actions taken by heterogeneous, self-interested firms.

It can certainly be argued that game-theoretic models of design and pricing have the potential to quantify policy outcomes that are currently ignored or discussed only in a qualitative way. As argued in Chapter 1, technology “configuration” is a prominent example of this. Single-stage game models of fuel economy choice and pricing are an initial application of game-theoretic principles to quantify how policy influences configuration. A simpler example is a pricing response to a fuel or sales tax increase. In part, tax increases may intend to sponsor fuel-efficient vehicle purchasing. However strategic firms can lower the prices of their vehicles in response to tax price

---

<sup>4</sup>The gradient condition above is a compact expression of this.



increases in order to maximize profits, potentially absorbing some of the desired response in vehicle purchases. Models that do not consider all the motivations that shape firm decisions cannot resolve these fairly intuitive aspects of market behavior.

The research presented in Chapters 2-5 and the future work proposed above primarily serve to make game-theoretic analyses of regulated differentiated product markets increasingly scientific and practical. Developing the underlying theory of equilibrium improves makes applications of game-theoretic models to policy questions more scientific. Advancing numerical methods for the computation of equilibrium has the potential to improve both how scientific an analysis is, as well as improving practicality.

Should these efforts succeed, the resulting analysis must still be informative and effectively communicated. Ensuring this requires deeper collaboration with other stakeholders in the policy debate, including automotive corporations, lobbyists, and key policy makers themselves. Engagement with these stakeholders will ensure that the primary questions of relevance to the policy debate are addressed by quantitative analyses, and that policy makers can act appropriately.

# Appendices

# Appendix A

## Mathematical Notation

*Sets.*  $\mathbb{N}$  denotes the natural numbers  $\{1, 2, \dots\}$ , and  $\mathbb{N}(N)$  denotes the natural numbers up to  $N$ , that is,  $\mathbb{N}(N) = \{1, \dots, N\}$ .  $\mathbb{R}$  denotes the set of real numbers  $(-\infty, \infty)$ ,  $[0, \infty)$  denotes the non-negative real numbers, and  $[0, \infty]$  denotes the extended non-negative half-line. We denote the  $(J - 1)$ -dimensional simplex  $\{(x_1, \dots, x_N) \in [0, 1]^N : \sum_{n=1}^N x_n = 1\}$  by  $\mathbb{S}(N)$ , and the  $J$ -dimensional “pyramid”  $\{(x_1, \dots, x_N) \in [0, 1]^N : \sum_{n=1}^N x_n \leq 1\}$  by  $\Delta(J)$ . Hyper-rectangles in  $\mathbb{R}^N$ , i.e. sets of the form  $[a_1, b_1] \times \dots \times [a_N, b_N]$  for some  $a_n, b_n \in \mathbb{R}$  with  $a_n < b_n$  for all  $n \in \mathbb{N}(N)$ , are denoted by  $[\mathbf{a}, \mathbf{b}]$  where  $\mathbf{a} = (a_1, \dots, a_N)$  and  $\mathbf{b} = (b_1, \dots, b_N)$ . For other sets, we typically use calligraphic upper case letters such as “ $\mathcal{A}$ ”. For any set  $\mathcal{A}$ ,  $|\mathcal{A}|$  denotes its cardinality. For any  $\mathcal{B} \subset \mathcal{A}$ ,  $\mathcal{A} \setminus \mathcal{B}$  denotes the set  $\{b \in \mathcal{A} : b \notin \mathcal{B}\}$ . For any set  $\mathcal{A}$ ,  $\mathfrak{F}(\mathcal{A})$  denotes the collection of finite subsets of  $\mathcal{A}$ .

*Symbols.* Bold, un-italicized symbols (e.g., “ $\mathbf{x}$ ”) denote vectors and matrices; typically we reserve lower case letters to refer to vectors and use upper case letters to refer to matrices; the vector of choice probabilities “ $\mathbf{P}$ ” is an exception made to conform with existing notation of these quantities. Throughout we use  $\mathbf{1}$  to denote a vector of ones of the appropriate size for the context in which it appears.  $\mathbf{I}$  always denotes the identity matrix of a size appropriate for the context. For any  $\mathbf{x} \in \mathbb{R}^N$ ,  $\text{diag}(\mathbf{x})$  denotes the  $N \times N$  diagonal matrix whose diagonal is  $\mathbf{x}$ . Any vector inequalities between vectors are to be taken componentwise: for example,  $\mathbf{x} < \mathbf{y}$  means  $x_n < y_n$  for all  $n$ . Random variables are denoted with capital letters “ $X$ ”, with random vectors being denoted with bold capital letters (e.g., “ $\mathbf{Q}$ ”). While this overlaps with our notation for matrices, it should not cause any confusion.  $\mathbb{P}$  denotes a probability and  $\mathbb{E}$  denotes an expectation. “log” always denotes the natural (base  $e$ ) logarithm. “ess sup” denotes the essential supremum of a measurable function, where the measure on measurable subsets of the domain should always be clear.

*Differentiation.* Our conventions for denoting differentiation follow [Munkres \(1991\)](#). We use the symbol “ $D$ ” to denote differentiation using subscripts to invoke additional specificity. Letting  $\mathbf{f} : \mathbb{R}^M \rightarrow \mathbb{R}^N$ ,  $(D_m f_n)(\mathbf{x})$  denotes the derivative of the  $n^{\text{th}}$  component function with respect to the  $m^{\text{th}}$  variable and  $(D\mathbf{f})(\mathbf{x})$  is the  $N \times M$  derivative matrix of  $\mathbf{f}$  at  $\mathbf{x}$  with components  $((D\mathbf{f})(\mathbf{x}))_{n,m} = (D_m f_n)(\mathbf{x})$ . Thus for  $f : \mathbb{R}^M \rightarrow \mathbb{R}$ ,  $(Df)(\mathbf{x})$  is a row vector. If  $f : \mathbb{R}^M \rightarrow \mathbb{R}$ , we define the gradient  $(\nabla f)(\mathbf{x}) \in \mathbb{R}^M$  as the transposed derivative:  $(\nabla f)(\mathbf{x}) = (Df)(\mathbf{x})^\top$ .

*Other Definitions.* Let  $\mathcal{X}$  be any topological space and let  $f : \mathcal{X} \rightarrow \mathbb{R}$ . We say  $\mathbf{x}^* \in \mathcal{X}$  is a *local maximizer (over  $\mathcal{X}$ )* of  $f$  if there exists a neighborhood of  $\mathbf{x}^*$ , say  $\mathcal{U}$ , such that  $f(\mathbf{x}^*) \geq f(\mathbf{x})$  for all  $\mathbf{x} \in \mathcal{U}$ . We say  $\mathbf{x}^* \in \mathcal{X}$  is a *maximizer (over  $\mathcal{X}$ )* of  $f$  if  $f(\mathbf{x}^*) \geq f(\mathbf{x})$  for all  $\mathbf{x} \in \mathcal{X}$ .

# Appendix B

## Examples for the Logit Model

### B.1 Demonstrations of the Basic Definitions

We first provide some examples of indirect utilities to illustrate properties (a-c). A linear in price utility, given by  $w(\mathbf{y}, p) = -\alpha(\mathbf{y})p$  for some  $\alpha : \mathcal{Y} \rightarrow (0, \infty)$ , satisfies (a-c). More generally, any “Cobb-Douglas” in price utility, given by  $w(\mathbf{y}, p) = -\alpha(\mathbf{y})p^{\beta(\mathbf{y})}$  with  $\alpha, \beta : \mathcal{Y} \rightarrow (0, \infty)$ , satisfies (a-c). A “Cobb-Douglas” specification for “remaining income,”  $w(\mathbf{y}, p) = \alpha(\mathbf{y})(\varsigma - p)^{\beta(\mathbf{y})}$  is a bit more complicated, being a function finite for all finite prices and satisfying (a-c) only for  $\beta : \mathcal{Y} \rightarrow (2\mathbb{N} + 1)$ , where  $2\mathbb{N} + 1$  denotes the set of odd positive integers: if  $\beta(\mathbf{y}) : \mathcal{Y} \rightarrow (-\infty, 0)$  then  $w$  is not finite for all finite  $p$ ; clearly  $w$  violates (a) if  $\beta(\mathbf{y}) = 0$ ; if  $\beta(\mathbf{y}) > 0$  is not an integer, then  $w$  is complex for  $p > \varsigma$ ; finally, if  $\beta(\mathbf{y}) \in \mathbb{N}$  is not an *odd* positive integer then  $w$  violates (a). The common “log-transformed” Cobb-Douglas in “remaining income” utility  $w(\mathbf{y}, p) = \alpha(\mathbf{y}) \log(\varsigma - p)$  for  $p < \varsigma < \infty$ ,  $\alpha : \mathcal{Y} \rightarrow (0, \infty)$ ,<sup>1</sup> is not finite for all finite prices. [Allenby and Rossi](#)’s negative log of price utility, given by  $w(\mathbf{y}, p) = -\alpha(\mathbf{y}) \log p$  for  $\alpha : \mathcal{Y} \rightarrow (0, \infty)$  satisfies (a-c) ([Allenby and Rossi, 1991](#)). Finally, the utility  $w(p) = -\alpha(\log p - \varepsilon \sin \log p)$ , where  $\alpha > 1$  and  $\varepsilon \in (0, 1)$ , satisfies (a-c).

We now demonstrate which of these utility functions is eventually log bounded and/or eventually decreases sufficiently quickly. Any linear in price or Cobb-Douglas in price utility is both eventually log bounded and eventually decreases sufficiently

---

<sup>1</sup>This log transformation usually occurs (see [Berry et al. \(1995\)](#), [Rossi et al. \(2006\)](#)) based on the observation that choices are invariant over increasing utility transformations, so that  $u'(\mathbf{y}, p) = e^{w(\mathbf{y}, p)} e^{v(\mathbf{y})}$  yields the same *random* choices as the specification introduced in the text, with the caveat that the *additive* errors introduced in the text are taken as *multiplicative* errors (with a related distribution) in the former specification. In a Cobb-Douglas specification for the former,  $u'(\mathbf{y}, p) \propto (\varsigma - p)^{\alpha(\mathbf{y})} = e^{\alpha(\mathbf{y}) \log(\varsigma - p)}$ , illustrating that the logarithm of this utility has the log-transformed specification for the price component.

quickly. For if  $\beta(\mathbf{y}) \geq 1$ ,  $(Dw)(\mathbf{y}, p) = -\alpha(\mathbf{y})\beta(\mathbf{y})p^{\beta(\mathbf{y})-1} \downarrow -\infty$  as  $p \rightarrow \infty$ . If  $\beta(\mathbf{y}) < 1$ , then although  $(Dw)(\mathbf{y}, p) = -\alpha(\mathbf{y})\beta(\mathbf{y})p^{-(1-\beta(\mathbf{y}))} \uparrow 0$  as  $p \rightarrow \infty$ ,

$$(Dw)(\mathbf{y}, p) - \frac{r}{p} = -\alpha(\mathbf{y})\beta(\mathbf{y})\frac{1}{p^{1-\beta(\mathbf{y})}} + \frac{r}{p} = \left(\frac{1}{p}\right) [r - \alpha(\mathbf{y})\beta(\mathbf{y})p^{\beta(\mathbf{y})}] \leq 0$$

if  $p \geq \sqrt[\beta(\mathbf{y})]{\alpha(\mathbf{y})\beta(\mathbf{y})/r}$  and hence  $w(\mathbf{y}, p)$  eventually decreases sufficiently quickly for any  $r$ . The class of negative log of price utility functions contain the most obvious examples of utilities that are neither eventually log bounded nor eventually decrease sufficiently quickly; particularly  $w(\mathbf{y}, p) \leq -\alpha(\mathbf{y}) \log p$  with  $\alpha(\mathbf{y}) \leq 1$ . If  $\alpha(\mathbf{y}) < 1$  there are no finite profit maximizing prices under this utility.

In the text we defined utilities with sub-quadratic second derivatives. Any linear in price utility has sub-quadratic second derivatives, since  $(D^2w)(\mathbf{y}, p) \equiv 0$ . More generally, under any Cobb-Douglas in price utility

$$\frac{(D^2w)(\mathbf{y}, p)}{(Dw)(\mathbf{y}, p)^2} = -\left(\frac{1}{\alpha(\mathbf{y})}\right) \left(\frac{\beta(\mathbf{y}) - 1}{\beta(\mathbf{y})}\right) \left(\frac{1}{p^{\beta(\mathbf{y})}}\right) = \left(\frac{\beta(\mathbf{y}) - 1}{\beta(\mathbf{y})}\right) \left(\frac{1}{w(\mathbf{y}, p)}\right),$$

and hence  $w$  has sub-quadratic second derivatives if  $\beta(\mathbf{y}) \geq 1$ . If  $\beta(\mathbf{y}) < 1$ , then  $w$  has sub-quadratic second derivatives only at  $(\mathbf{y}, p)$  such that  $|w(\mathbf{y}, p)| > (1 - \beta(\mathbf{y}))/\beta(\mathbf{y})$ , i.e.  $p > \sqrt[\beta(\mathbf{y})]{(1 - \beta(\mathbf{y}))/(\alpha(\mathbf{y})\beta(\mathbf{y}))}$ . Finally, if  $w(\mathbf{y}, p) = -\alpha(\mathbf{y}) \log p$  then  $(D^2w)(\mathbf{y}, p)/(Dw)(\mathbf{y}, p)^2 \equiv 1/\alpha(\mathbf{y})$  and hence  $w$  has sub-quadratic second derivatives if  $\alpha(\mathbf{y}) > 1$ . Hence far from requiring concavity, some convex utility functions have sub-quadratic second derivatives.

## B.2 Examples of the $\zeta$ map

Let  $\alpha(\mathbf{y}) \equiv \alpha > 0$ . For the linear-in-price utility,  $\zeta_j(\mathbf{p}) = \hat{\pi}_f(\mathbf{p}) + 1/\alpha$  with the fixed-point equation being  $p_j = c_j + \hat{\pi}_f(\mathbf{p}) + 1/\alpha$ . For any Cobb-Douglas in price utility,  $\zeta_j(\mathbf{p}) = \hat{\pi}_f(\mathbf{p}) + (1/(\alpha\beta))p_j^{1-\beta}$  with the fixed-point equation being  $p_j = c_j + \hat{\pi}_f(\mathbf{p}) + (1/(\alpha\beta))p_j^{1-\beta}$ . For negative log of price,  $\zeta_j(\mathbf{p}) = \hat{\pi}_f(\mathbf{p}) + (1/\alpha)p_j$  with the fixed-point equation being  $p_j = c_j + \hat{\pi}_f(\mathbf{p}) + (1/\alpha)p_j$ .

Our proof that the negative log of price utility has no finite profit maximizing prices can be strengthened using the relationship between  $\zeta$  and the profit gradients. We already know that  $w(\mathbf{y}, p)/1 = -(\alpha(\mathbf{y})/1) \log p$  does not eventually decrease sufficiently quickly when  $\alpha(\mathbf{y}) \leq 1$ . We have also observed that  $\zeta_j(\mathbf{p}) = \hat{\pi}_f(\mathbf{p}) + (1/\alpha_j)p_j$ ,

which implies that

$$\zeta_j(\mathbf{p}) - (p_j - c_j) = (\hat{\pi}_f(\mathbf{p}) + c_j) + \left(\frac{1}{\alpha_j} - 1\right) p_j = (\hat{\pi}_f(\mathbf{p}) + c_j) + \left(\frac{1 - \alpha_j}{\alpha_j}\right) p_j.$$

Thus, if  $\alpha_j = \alpha(\mathbf{y}_j) \leq 1$ , the  $j^{\text{th}}$  price derivative of profit is always positive. While we have already shown that only infinite prices maximize profits under this utility when  $\alpha_j < 1$ , this shows the same holds for  $\alpha_j = 1$  as well even though the corresponding maximal profits are finite.

We now present an example of a utility function for which has finite profit-maximizing prices but for which a “local” criterion restricting profit maximization at infinity fails. This local criterion is simply that profits decrease for all sufficiently large prices. Let  $w(p) = -\alpha(\log p - \varepsilon \sin \log p)$  with  $\alpha > 1$  and  $\varepsilon \in [1 - \alpha^{-1}, 1)$ . Then  $p_j - c_j - \zeta_j(\mathbf{p}) \geq 0$  if and only if

$$p_j \left(1 - \frac{1}{\alpha(1 - \varepsilon \cos \log p_j)}\right) \geq c_j + \hat{\pi}_f(\mathbf{p}). \quad (\text{B.1})$$

But based on our choice of  $\varepsilon$ , there exist arbitrarily large  $p_j$  such that the left hand side above is non-positive: For all  $\bar{p}$  there exists some  $p_j > \bar{p}$  such that  $\alpha(1 - \varepsilon \cos \log p_j) = \alpha(1 - \varepsilon) \leq 1$ , which implies the claim. Since  $c_j + \hat{\pi}_f(\mathbf{p})$  is positive (or rather is for all  $\mathbf{p}$  that matter), the inequality (B.1) is violated and there exist arbitrarily large  $p_j$  such that profits *increase*, locally, with  $p_j$ , despite the fact that profits must vanish as  $\mathbf{p}_f \rightarrow \infty$  since this utility is eventually log bounded. That is, the *local* criterion for finite profit maximizing prices is violated.

# Appendix C

## An Extended Discussion of Numerical Methods

In this section we extend our discussion of the fixed-point iteration and Newton's method.

All of the methods we implement can be built to ignore products with excessively low choice probabilities. That is, one can ignore price updates for all products with  $P_j(\mathbf{p}) \leq \varepsilon_P$ , where  $\varepsilon_P$  is some small value (say  $10^{-10}$ ). Products with a choice probability this small (or smaller) need not be considered a part of the market in the price equilibrium computations. For example, [Wards \(2004-2007\)](#) reports total sales of cars and light trucks during 2005 as  $N = 16,947,754$ .<sup>1</sup> Because expected demand is defined by  $\mathbb{E}[Q_j(\mathbf{p})] = NP_j(\mathbf{p})$ , any  $\varepsilon_P \leq 0.5 * N^{-1} \approx 3 \times 10^{-8}$  ignores any vehicle that, as priced, is not *expected* to have a *single* customer out of the millions of customers that bought or considered buying new vehicles. There are also technical reasons for this truncation. Particularly,  $\mathbf{\Lambda}(\mathbf{p})$  and  $(D\tilde{\nabla}\hat{\pi})(\mathbf{p})$  become singular as  $P_j(\mathbf{p}) \rightarrow 0$ , for any  $j$ . Truncating avoids this non-singularity and hopefully helps conditioning.

### C.1 Our Fixed-Point Iteration

The only potential complication that may arise in iterating

$$\mathbf{p} \leftarrow \mathbf{c} + \mathbf{\Lambda}(\mathbf{p})^{-1}(\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top(\mathbf{p} - \mathbf{c}) - \mathbf{P}(\mathbf{p})).$$

is that  $\lambda_j(\mathbf{p}) \rightarrow 0$  for some  $j$ , as suggested above. This happens if  $p_j \rightarrow \varsigma_*$ , which will be finite for the finite-sample models applied in computations. Technically,  $\zeta_j(\mathbf{p})$  is

---

<sup>1</sup>Particularly, 7,667,066 cars and 9,280,688 light trucks.



not defined when  $\lambda_j(\mathbf{p}) = 0$ . However it is also true that  $(D_j \hat{\pi}_f)(\mathbf{p}) = 0$ . Hence the objective of the iteration, to zero the price derivatives of profit, has been achieved for the  $j^{\text{th}}$  price. Furthermore, so long as  $\lambda_j(\mathbf{p}) = 0$ , the value of  $p_j$  does not determine the value of the fixed-point step for any other prices. These facts are guaranteed by the following proposition.

**Proposition C.1.1.** *Under assumption (a)  $P_j(\mathbf{p})$  and  $\lambda_j(\mathbf{p})$  are either both zero or neither is. If either is, then also  $\gamma_{j,k}(\mathbf{p}) = \gamma_{k,j}(\mathbf{p}) = 0$  for all  $k \in \mathbb{N}(J)$ .*

## C.2 Newton’s Method

### C.2.1 Computing Jacobian Matrices

Standard “exact” or quasi-Newton methods either always or periodically require the Jacobian matrix  $(D\mathbf{F})(\mathbf{x})$ . Using finite differences to approximate either Jacobian matrix increases the computational burden by an order of magnitude by requiring work equivalent to  $J$  evaluations of the function  $\mathbf{F}$ . In our 993 vehicle example, approximating  $(D\mathbf{F})(\mathbf{x})$  *once* with finite differences would take roughly 993 evaluations of the combined gradient, when the work of less than 50 evaluations appears to sufficient to converge to equilibrium prices using the  $\zeta$  fixed-point iteration. We recommend directly approximating  $(D\mathbf{F})(\mathbf{x})$  using integral expressions for  $(D\tilde{\nabla}\hat{\pi})(\mathbf{p})$ ,  $(D\boldsymbol{\eta})(\mathbf{p})$ , and  $(D\zeta)(\mathbf{p})$ . In this subsection we report formulae that enable this task. Another alternative is to use automatic differentiation, but we are skeptical that this would in fact be faster than the direct formulae provided here.

#### Combined Gradient

Assuming a second application of the Leibniz Rule holds, we can derive integral expressions for the second derivatives  $(D_l D_k \hat{\pi}_{f(k)})(\mathbf{p})$  through

$$((D\tilde{\nabla}\hat{\pi})(\mathbf{p}))_{k,l} = (D_l D_k \hat{\pi}_{f(k)})(\mathbf{p}) = \int (D_l D_k \hat{\pi}_{f(k)}^L)(\boldsymbol{\theta}, \mathbf{p}) d\mu(\boldsymbol{\theta}).$$

**Proposition C.2.1.** *Let  $w$  be twice continuously differentiable in  $p$  and suppose a second application of the Leibniz Rule holds for the Mixed Logit choice probabilities at*

**p.** Set

$$\begin{aligned}\phi_{k,l}(\mathbf{p}) &= \int (Dw_k)(\boldsymbol{\theta}, p_k) P_k^L(\boldsymbol{\theta}, \mathbf{p}) P_l^L(\boldsymbol{\theta}, \mathbf{p}) (Dw_l)(\boldsymbol{\theta}, p_l) d\mu(\boldsymbol{\theta}) \\ \psi_{k,l}(\mathbf{p}) &= \int (Dw_k)(\boldsymbol{\theta}, p_k) P_k^L(\boldsymbol{\theta}, \mathbf{p}) \hat{\pi}_{f(k)}^L(\boldsymbol{\theta}, \mathbf{p}) P_l^L(\boldsymbol{\theta}, \mathbf{p}) (Dw_l)(\boldsymbol{\theta}, p_l) d\mu(\boldsymbol{\theta}) \\ \chi_k(\mathbf{p}) &= \left(\frac{1}{2}\right) \int ((D^2w_k)(\boldsymbol{\theta}, p_k) + (Dw_k)(\boldsymbol{\theta}, p_k)^2) \\ &\quad \times P_k^L(\boldsymbol{\theta}, \mathbf{p}) ((p_k - c_k) - \hat{\pi}_{f(k)}^L(\boldsymbol{\theta}, \mathbf{p})) d\mu(\boldsymbol{\theta})\end{aligned}$$

(i) Component form: *Setting*

$$\xi_{k,l}(\mathbf{p}) = \delta_{k,l}(\lambda_k(\mathbf{p}) + \chi_k(\mathbf{p})) - \gamma_{k,l}(\mathbf{p}) - (p_k - c_k)\varphi_{k,l}(\mathbf{p})$$

we have

$$(D_l D_k \hat{\pi}_{f(k)})(\mathbf{p}) = \xi_{k,l}(\mathbf{p}) + 2\psi_{k,l}(\mathbf{p}) + \delta_{f(k),f(l)} \xi_{l,k}(\mathbf{p})$$

(ii) Matrix form: Let  $\Phi(\mathbf{p})$ ,  $\Psi(\mathbf{p})$  and  $\mathbf{X}(\mathbf{p}) = \text{diag}(\boldsymbol{\chi}(\mathbf{p}))$  be the matrices of these quantities. Also set

$$\Xi(\mathbf{p}) = \Lambda(\mathbf{p}) - \Gamma(\mathbf{p}) - \text{diag}(\mathbf{p} - \mathbf{c})\Phi(\mathbf{p}) + \mathbf{X}(\mathbf{p}).$$

and

$$(\tilde{\Xi}(\mathbf{p}))_{k,l} = \begin{cases} \xi_{k,l}(\mathbf{p}) & \text{if } f(k) = f(l) \\ 0 & \text{if } f(k) \neq f(l) \end{cases}$$

Then

$$(D\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \Xi(\mathbf{p}) + 2\Psi(\mathbf{p}) + \tilde{\Xi}(\mathbf{p})^\top. \quad (\text{C.1})$$

*Proof.* Proof: To see that this only relies on a second application of the Leibniz Rule to the *choice probabilities*, note that

$$(D_l D_k \hat{\pi}_{f(k)})(\mathbf{p}) = \sum_{j \in \mathcal{J}_{f(k)}} (D_l D_k P_j)(\mathbf{p})(p_j - c_j) + \delta_{f(k),f(l)} (D_k P_l)(\mathbf{p}) + (D_l P_k)(\mathbf{p})$$

and thus the continuous second-order differentiability of  $\hat{\pi}_f(\mathbf{p})$  depends only on the second-order continuous differentiability of  $\mathbf{P}_f$ . This result is then an immediate consequence of the validity of the Leibniz Rule and Eqn. (D.2).

The validity of a second application of the Leibniz Rule to the choice probabilities is ensured by the following condition.

**Proposition C.2.2.** *Let  $(u, \vartheta, \mu) = (w + v, \vartheta, \mu)$  be such that*

- (i)  $w(\boldsymbol{\theta}, \mathbf{y}, \cdot) : (0, \varsigma_*) \rightarrow \mathbb{R}$  *is twice continuously differentiable for all  $\mathbf{y} \in \mathcal{Y}$  and  $\mu$ -a.e.  $\boldsymbol{\theta} \in \mathcal{T}$*
- (ii) *for all  $(\mathbf{y}, p) \in \mathcal{Y} \times (0, \varsigma_*)$ ,  $|(D^2w)(\cdot, \mathbf{y}, q) + (Dw)(\cdot, \mathbf{y}, q)^2| e^{u(\cdot, \mathbf{y}, q) - \vartheta(\cdot)} : \mathcal{T} \rightarrow [0, \infty)$  is uniformly  $\mu$ -integrable for all  $q$  in some neighborhood of  $p$ .*
- (iii) *for all  $(\mathbf{y}, p), (\mathbf{y}', p') \in \mathcal{Y} \times (0, \varsigma_*)$ ,*

$$|(Dw)(\cdot, \mathbf{y}, q)| e^{u(\cdot, \mathbf{y}, q) - \vartheta(\cdot)} e^{u(\cdot, \mathbf{y}', q') - \vartheta(\cdot)} |(Dw)(\cdot, \mathbf{y}', q')| : \mathcal{T} \rightarrow [0, \infty)$$

*is uniformly  $\mu$ -integrable for all  $(q, q')$  in some neighborhood of  $(p, p')$ .*

*Then a second application of the Leibniz Rule holds for the Mixed Logit choice probabilities, which are also continuously differentiable on  $(\mathbf{0}, \varsigma_* \mathbf{1})$ .*

This is proved in the same manner as Proposition 4.1.2.

We also observe the following.

**Proposition C.2.3.** *If  $P_k(\mathbf{p}) = 0$  then  $(D_l D_k \hat{\pi}_{f(k)})(\mathbf{p}) = (D_k D_l \hat{\pi}_{f(l)})(\mathbf{p}) = 0$  for all  $l \in \mathbb{N}(J)$ .*

The proof follows from the derivative formulae given above. Of course, if  $P_k(\mathbf{p}) = 0$  then  $(D_k \hat{\pi}_{f(k)})(\mathbf{p}) = 0$  as well and we have a similar situation as with the fixed-point iteration: (i) the Newton system is consistent for any  $s_k^N(\mathbf{p}) \in \mathbb{R}$  and (ii)  $s_l^N(\mathbf{p})$  does not depend on  $s_k^N(\mathbf{p})$  for all  $l \in \mathbb{N}(J) \setminus \{k\}$ . Thus, in practice we restrict attention to the Newton step defined by the submatrix of  $(D\tilde{\nabla}\hat{\pi})(\mathbf{p})$  formed by rows and columns indexed by  $\{j : P_j(\mathbf{p}) > \varepsilon_P\}$ .

The formulae above give the following expression of the profit Hessians.

**Corollary C.2.4.** *Let  $w$  be twice continuously differentiable in  $p$  and suppose a second application of the Leibniz Rule holds for the Mixed Logit choice probabilities. Firm  $f$ 's profit Hessian is given by*

$$(D_f \nabla_f \hat{\pi}_f)(\mathbf{p}) = \Xi_{f,f}(\mathbf{p}) + 2\Psi_{f,f}(\mathbf{p}) + \Xi_{f,f}(\mathbf{p})^\top.$$

**The  $\eta$  map.**

Applying  $\mathbf{F}(\mathbf{p}) = \mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p})$ , we have  $(D\mathbf{F})(\mathbf{p}) = \mathbf{I} - (D\boldsymbol{\eta})(\mathbf{p})$ .  $(D\boldsymbol{\eta})(\mathbf{p})$  solves the linear matrix equation  $(\tilde{D}\mathbf{P})(\mathbf{p})^\top (D\boldsymbol{\eta})(\mathbf{p}) = -(\mathbf{A}(\mathbf{p}) + (D\mathbf{P})(\mathbf{p}))$  where  $(\mathbf{A}(\mathbf{p}))_{k,l} = \sum_{j \in \mathcal{J}_{f(k)}} (D_l D_k P_j)(\mathbf{p}) \eta_j(\mathbf{p})$ . This is easily derived from the defining formula  $(\tilde{D}\mathbf{P})(\mathbf{p})^\top \boldsymbol{\eta}(\mathbf{p}) = -\mathbf{P}(\mathbf{p})$ .

### The $\zeta$ map.

Applying  $\mathbf{F}(\mathbf{p}) = \mathbf{p} - \mathbf{c} - \zeta(\mathbf{p})$ , we have  $(D\mathbf{F})(\mathbf{p}) = \mathbf{I} - (D\zeta)(\mathbf{p})$ .  $(D\zeta)(\mathbf{p})$  can be computed using the following formula.

**Lemma C.2.5.**

$$(D_l \zeta_k) = \lambda_k^{-1} \left[ \delta_{k,l} \left[ \int P_k^L ((D^2 w_k) + (D w_k)^2) (\hat{\pi}_{f(k)}^L - \zeta_k) - \lambda_k \right] \right. \\ \left. + \zeta_k (\phi_{k,l} + \gamma_{k,l}) + \delta_{f(k),f(l)} \phi_{k,l} (p_l - c_l) + \gamma_{l,k} - 2\psi_{k,l} \right]$$

### C.2.2 Quasi-Newton Methods and the BFGS Update.

Quasi-Newton methods build and apply approximations to the Jacobian matrices  $(D\mathbf{F})(\mathbf{p})$  to reduce the often large burden of their computation (Dennis and Schnabel, 1996). These methods generalize Newton’s method by taking steps  $\mathbf{s}^{QN}$  that satisfy  $\mathbf{A}\mathbf{s}^{QN} = -\mathbf{F}(\mathbf{p})$  for some (nonsingular and well-conditioned) matrix  $\mathbf{A}$ . The price update then takes the form  $\mathbf{p} \leftarrow \mathbf{p} + \mathbf{s}^{QN}$ . For example, Newton’s method results from setting  $\mathbf{A} = (D\mathbf{F})(\mathbf{p})$ . The fixed-point iteration  $\mathbf{p} \leftarrow \mathbf{c} + \zeta(\mathbf{p})$  can also be considered a Quasi-Newton method for  $\mathbf{F} = (\tilde{\nabla}\hat{\pi})$  by setting  $\mathbf{A} = \mathbf{\Lambda}(\mathbf{p})$ .<sup>2</sup>

The most prominent quasi-newton method for nonlinear equations is probably the *BFGS update* (Broyden et al., 1973; Dennis and Schnabel, 1996). Taking a BFGS quasi-Newton step will always be strictly more burdensome than taking a fixed-point step because either the combined gradient,  $\boldsymbol{\eta}$ , or  $\zeta$  itself remain on the right-hand side. Outside of initialization (and periodic re-initialization) of  $\mathbf{A}$  with the actual Jacobian matrix  $(D\mathbf{F})(\mathbf{p})$  and the computation of its QR factors, directly updating the QR factorization of  $\mathbf{A}$  and solving for  $\mathbf{s}^{QN}$  requires only  $\mathcal{O}(J^2)$  flops (Dennis and Schnabel, 1996). Some globalization strategy is generally needed to ensure convergence and further compounds the computational effort required. Line search, the hookstep, and the dogleg step are easily implemented by exchanging all occurrences of the Jacobian  $(D\mathbf{F})(\mathbf{p})$  with its Quasi-Newton approximation  $\mathbf{A}$ .

---

<sup>2</sup>In fact, this could lead to a proof of the local linear convergence of the fixed-point iteration if  $\mathbf{\Lambda}(\mathbf{p})$  can be shown to be an approximation to  $(D\tilde{\nabla}\hat{\pi})(\mathbf{p})$  of “bounded deterioration” in some neighborhood of equilibrium; see (Broyden et al., 1973).

### C.2.3 Inexact Newton Methods

Because Inexact Newton methods are not as widely known, we present a relatively detailed discussion of these methods and our implementation here.

As stated in the Chapter 4, inexact Newton steps are simply “inexact” solutions to the Newton system; that is, an inexact Newton step  $\mathbf{s}^{IN}$  is any vector that satisfies

$$\|\mathbf{F}(\mathbf{x}) + (D\mathbf{F})(\mathbf{x})\mathbf{s}^{IN}\| \leq \delta \|\mathbf{F}(\mathbf{x})\| \quad (\text{C.2})$$

for some fixed  $\delta \in (0, 1)$ . In principle, one could implement an inexact Newton method on a problem with few variables using direct methods of solution. In fact, if in quasi-Newton fashion  $\mathbf{s}^{IN}$  is defined as the solution to  $\mathbf{A}\mathbf{s}^{IN} = -\mathbf{F}(\mathbf{x})$  for some  $\mathbf{A}$  satisfying  $\|\mathbf{I} - (D\mathbf{F})(\mathbf{x})\mathbf{A}^{-1}\| \leq \delta$ , then the inexact Newton condition holds. In this sense, some quasi-Newton methods could be considered inexact Newton methods.<sup>3</sup> By appropriately choosing a sequence of  $\delta$ 's, the local asymptotic convergence rate of a sequence generated by an inexact Newton's method can be made fully q-quadratic (Dembo et al., 1982; Eisenstat and Walker, 1994). Of course, this will also require increasingly burdensome computations of inexact Newton steps that satisfy increasingly strict inexact Newton conditions.

When using iterative linear solvers like GMRES to approximately solve  $(D\mathbf{F})(\mathbf{x})\mathbf{s}^{IN} = -\mathbf{F}(\mathbf{x})$ , the inexact Newton condition is a statement on the relative residual of the approximate solution to the Newton system. In this context this approach may be called a “truncated” Newton method. It is this area that is relevant for large-scale problems and on which we focus.

#### Directional Derivatives

When implementing an iterative linear system solver like GMRES only matrix-vector products of the type  $(D\mathbf{F})(\mathbf{x})\mathbf{s}$  will be required. These directional derivatives can be approximated with “directional” finite differences in much less work than it takes to approximate the entire Jacobian matrix. For example, the first-order formula

$$(D\mathbf{F})(\mathbf{x})\mathbf{s} \approx h^{-1}(\mathbf{F}(\mathbf{x} + h\mathbf{s}) - \mathbf{F}(\mathbf{x})),$$

---

<sup>3</sup>However if  $\delta \rightarrow 0$  during an iteration, this condition would in some sense require that  $\mathbf{A} \rightarrow (D\mathbf{F})$ . This is *not* true for the most popular and successful quasi-Newton methods like the BFGS update that do have q-superlinear locally asymptotic convergence rates (Broyden et al., 1973; Dennis and Schnabel, 1996). On the other hand,  $\delta$  can be chosen to be a constant if a q-linear locally asymptotic convergence rate is suitable (Pernice and Walker, 1998).

requires only a single additional evaluation of  $\mathbf{F}$  per approximate evaluation of  $(D\mathbf{F})(\mathbf{x})\mathbf{s}$ . Higher-order formulae requiring 2 and 4 additional evaluations of  $\mathbf{F}$  are easy to derive; see [Pernice and Walker \(1998\)](#). Since this is repeated *at each step* of iterative linear solvers, *each step of an iterative Newton system solver using directional finite differences is at least as expensive as a  $\zeta$  fixed-point step*. If an iterative solver should take 100 steps to compute an inexact Newton step having small enough residual to satisfy the inexact Newton condition, then we could have equivalently taken 100, 200, and 400  $\zeta$  fixed-point steps with the first, second, and fourth order formulae available in [Pernice and Walker \(1998\)](#). In our examples, this is easily enough to converge to equilibrium prices from relatively arbitrary initial conditions.<sup>4</sup> While we have found appropriately implemented iterative linear solvers take tens, rather than hundreds, of iterations to solve the inexact Newton system, this comparison highlights the increase in work implicit even in this “fast” version of Newton’s method.

Generally speaking, whether or not is more efficient to compute the directional finite differences or to compute the Jacobian and carry out the actual multiplications depends on how many steps the iterative solver requires to satisfy the inexact Newton condition. Computing and using the Jacobian matrix is an  $\mathcal{O}((S+N)J^2)$  process while using directional finite differences is an  $\mathcal{O}(SN \sum_{f=1}^F J_f^2)$  process when it takes  $N \in \mathbb{N}$  steps to satisfy the inexact Newton condition. To leading order, it is worthwhile to use the full Jacobian when

$$(S+N)J^2 \leq SN \sum_{f=1}^F J_f^2. \quad (\text{C.3})$$

If  $S > J^2 / \sum_{f=1}^F J_f^2$  (which should hold, generally speaking, because  $J^2 / \sum_{f=1}^F J_f^2 \leq F$  and  $S$  should probably be larger than  $F$ ), then Eqn. (C.3) is equivalent to

$$N \geq S \left( \frac{J^2 / \sum_{f=1}^F J_f^2}{S - J^2 / \sum_{f=1}^F J_f^2} \right) \quad (\text{C.4})$$

Moreover

$$S \left( \frac{J^2 / \sum_{f=1}^F J_f^2}{S - J^2 / \sum_{f=1}^F J_f^2} \right) \geq \frac{J^2}{\sum_{f=1}^F J_f^2}$$

---

<sup>4</sup>In their implementation of the **GMRES** method in the context of an inexact Newton method, [Pernice and Walker \(1998\)](#) only use higher order finite-differencing formulas at restarts. The first-order approach must still be applied at each step of the iterative solver.

and

$$\lim_{S \uparrow \infty} \left[ S \left( \frac{J^2 / \sum_{f=1}^F J_f^2}{S - J^2 / \sum_{f=1}^F J_f^2} \right) \right] = \frac{J^2}{\sum_{f=1}^F J_f^2}.$$

Thus if **GMRES** (or other similar iterative numerical method) takes  $N > J^2 / \sum_{f=1}^F J_f^2$  steps, computing using the Jacobian matrix is asymptotically more efficient as  $S \rightarrow \infty$ . In our 5,298 vehicle example,  $J^2 / \sum_{f=1}^F J_f^2 \approx 12$  while in the 993 vehicle example  $J^2 / \sum_{f=1}^F J_f^2 \approx 24$ ; in either case  $J^2 / \sum_{f=1}^F J_f^2 \leq F = 38$ . As we shall see below, our preconditioned version of **GMRES** for  $(\tilde{\nabla}\hat{\pi})$  takes roughly this many steps, and thus we feel justified in using the full Jacobian in our calculations. For the fixed-point formulations, **GMRES** generally converges rapidly ( $\leq 10$  steps). Direct comparison of calculations done using finite directional differences and the full Jacobians illustrate that using the Jacobians is more efficient.

### Preconditioning

As is well known, preconditioning is key to the effectiveness of iterative linear solvers; see [Golub and Loan \(1996\)](#). We have not found  $\mathbf{F}(\mathbf{x}) = \mathbf{x} - \mathbf{c} + \boldsymbol{\eta}(\mathbf{x})$  or  $\mathbf{F}(\mathbf{x}) = \mathbf{x} - \mathbf{c} + \boldsymbol{\zeta}(\mathbf{x})$  to need preconditioning. However, when  $\mathbf{F}(\mathbf{x}) = (\tilde{\nabla}\hat{\pi})(\mathbf{x})$  we have found the preconditioned system

$$\boldsymbol{\Lambda}(\mathbf{p})^{-1}(D\tilde{\nabla}\hat{\pi})(\mathbf{p})\mathbf{s}^{IN} = -\boldsymbol{\Lambda}(\mathbf{p})^{-1}(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \mathbf{c} + \boldsymbol{\zeta}(\mathbf{p}) - \mathbf{p} \quad (\text{C.5})$$

to be very rapidly solved for reasonably small relative errors. The following result, an extension of the ideas in Chapter 3, Section 3.1, relates the Jacobian of the combined gradient to the Jacobians of the fixed-point maps in equilibrium.

**Lemma C.2.6.**  $\mathbf{I} - (D\boldsymbol{\zeta})(\mathbf{p}) = \boldsymbol{\Lambda}(\mathbf{p})^{-1}(D\tilde{\nabla}\hat{\pi})(\mathbf{p})$  for any simultaneously stationary  $\mathbf{p}$ .

*Proof.* This follows from differentiating  $(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \boldsymbol{\Lambda}(\mathbf{p})(\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p}))$  via the product rule, recognizing that  $\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p}) = \mathbf{0}$  in equilibrium and  $D[\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p})] = \mathbf{I} - (D\boldsymbol{\zeta})(\mathbf{p})$ .  $\square$

In other words, Newton-type methods applied to  $\mathbf{F}(\mathbf{x}) = \mathbf{x} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{x})$  and  $\mathbf{F}(\mathbf{x}) = (\tilde{\nabla}\hat{\pi})(\mathbf{x})$  preconditioned as above end up being essentially the same iterations, close enough to equilibrium.

Note that **GMRES**, used successfully on this preconditioned system Eqn. (C.5), will

ensure that

$$\|\mathbf{\Lambda}(\mathbf{p})^{-1}(\tilde{\nabla}\hat{\pi})(\mathbf{p}) + \mathbf{\Lambda}(\mathbf{p})^{-1}(D\tilde{\nabla}\hat{\pi})(\mathbf{p})\mathbf{s}^{IN}\| \leq \delta' \|\mathbf{\Lambda}(\mathbf{p})^{-1}(\tilde{\nabla}\hat{\pi})(\mathbf{p})\| \quad (\text{C.6})$$

for some  $\delta'$ . This is distinct from the inexact Newton condition Eqn. (C.2). The following proposition gives modified tolerances for the preconditioned system to ensure satisfaction of the original system.

**Proposition C.2.7.** *Let  $\delta > 0$  be given. If Eqn. (C.6) is satisfied with  $\delta'(\mathbf{p}, \delta) \leq \delta$  given by*

$$\delta'(\mathbf{p}, \delta) = \left( \frac{\|(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_2}{\max_j \{|\lambda_j(\mathbf{p})|\} \|\mathbf{\Lambda}(\mathbf{p})^{-1}(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_2} \right) \delta, \quad (\text{C.7})$$

then Eqn. (C.2) is satisfied.

This is a consequence of the following general result, which we state without proof.

**Lemma C.2.8.** *Let  $\mathbf{b} \in \mathbb{R}^N$  and  $\mathbf{A}, \mathbf{M} \in \mathbb{R}^{N \times N}$  be nonsingular. Then*

$$\frac{\|\mathbf{Ax} - \mathbf{b}\|}{\|\mathbf{b}\|} \leq \alpha \left( \frac{\|\mathbf{M}^{-1}\mathbf{Ax} - \mathbf{M}^{-1}\mathbf{b}\|}{\|\mathbf{M}^{-1}\mathbf{b}\|} \right) \quad (\text{C.8})$$

where  $\alpha \in [1, \kappa(\mathbf{M})]$  is given by

$$\alpha = \frac{\|\mathbf{M}\| \|\mathbf{M}^{-1}\mathbf{b}\|}{\|\mathbf{b}\|} = \|\mathbf{M}\| \|\mathbf{M}^{-1} \left( \frac{\mathbf{b}}{\|\mathbf{b}\|} \right)\|.$$

This implies that

$$\frac{\|\mathbf{Ax} - \mathbf{b}\|}{\|\mathbf{b}\|} \leq \delta \quad \text{if} \quad \frac{\|\mathbf{M}^{-1}\mathbf{Ax} - \mathbf{M}^{-1}\mathbf{b}\|}{\|\mathbf{M}^{-1}\mathbf{b}\|} \leq \frac{\delta}{\alpha}.$$

Note that the preconditioned system must always be solved to a *stricter* tolerance than is desired for the un-preconditioned system using this bound. Additionally, computing  $\alpha$  for a generic preconditioner  $\mathbf{M}$  relies on the ability to compute  $\|\mathbf{M}\|$ .

Eqn. (C.7) simply adopts the 2-norm and applies the formula (Golub and Loan, 1996)

$$\|\mathbf{\Lambda}(\mathbf{p})\|_2 = \sqrt{\max_j \{|\lambda_j(\mathbf{p})|^2\}} = \max_j \{|\lambda_j(\mathbf{p})|\}$$

In addition,  $\delta'(\mathbf{p}, \delta) \leq \delta$ , simply because

$$\frac{\|(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_2}{\|\mathbf{\Lambda}(\mathbf{p})\| \|\mathbf{\Lambda}(\mathbf{p})^{-1}(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_2} = \frac{\|\mathbf{\Lambda}(\mathbf{p})(\mathbf{p} - \mathbf{c} - \zeta(\mathbf{p}))\|_2}{\|\mathbf{\Lambda}(\mathbf{p})\| \|\mathbf{p} - \mathbf{c} - \zeta(\mathbf{p})\|_2} \leq 1.$$



Eqn. (C.8) also implies that if Eqn. (C.6) holds with  $\delta' > 0$ , then

$$\frac{\|(\tilde{\nabla}\hat{\pi})(\mathbf{p}) + (D\tilde{\nabla}\hat{\pi})(\mathbf{p})\mathbf{s}^{IN}\|_2}{\|(\tilde{\nabla}\hat{\pi})(\mathbf{p})\|_2} \leq \kappa_2(\mathbf{\Lambda}(\mathbf{p}))\delta'$$

where  $\kappa_2(\mathbf{\Lambda}(\mathbf{p})) = \|\mathbf{\Lambda}(\mathbf{p})\|_2\|\mathbf{\Lambda}(\mathbf{p})^{-1}\|_2$  is the (2-norm) condition number of  $\mathbf{\Lambda}(\mathbf{p})$ . This equation, while the more compact representation, can also be overly conservative as clearly illustrated in Fig. C.1. It is unlikely that  $\kappa(\mathbf{\Lambda}(\mathbf{p}))$  is a *tight* upper bound on the multiplier in Eqn. (C.7). In fact, the multiplier on  $\delta$  depends only on the norm of  $\mathbf{\Lambda}(\mathbf{p})^{-1}\mathbf{x}$  at a single point on the surface of the unit sphere in  $\mathbb{R}^J$  rather than  $\|\mathbf{\Lambda}(\mathbf{p})^{-1}\|_2$ , the maximum norm of  $\mathbf{\Lambda}(\mathbf{p})^{-1}\mathbf{x}$  over this entire sphere. Our examples in Fig. C.1 bear this out, having condition numbers many orders of magnitude larger than the multiplier in Eqn. (C.7).

The power of the preconditioning is that the preconditioned system Eqn. (C.6) appears to be solved to a relative error of  $\delta'(\mathbf{p}, \delta)$  *much* faster than the original system can be solved to a relative error of  $\delta$ , even though  $\delta'(\mathbf{p}, \delta) \leq \delta$ . As can be seen in Fig. C.1, solving the preconditioned system to  $\delta'(\mathbf{p}, \delta)$  can achieve a relative error in the original system below  $\delta = 10^{-10}$  in roughly *four orders of magnitude fewer iterations* than solving the original system to this same relative error for prices near equilibrium. Away from equilibrium, GMRES may not be able to solve the original system to small relative errors like  $10^{-6}$  at all. Thus using the original system would appear to slow, if not halt, an implementation of the inexact Newton's method.

## GMRES Hookstep

We describe an implementation of the locally constrained optimal or “hookstep” (Dennis and Schnabel, 1996) modification suitable for GMRES as first suggested by Viswanath (2007, 2008). First, we recall the basic structure of model trust region methods; see (Dennis and Schnabel, 1996, Chapter 6, Section 4). We then adopt this structure to the case of Krylov subspace methods, particularly GMRES.

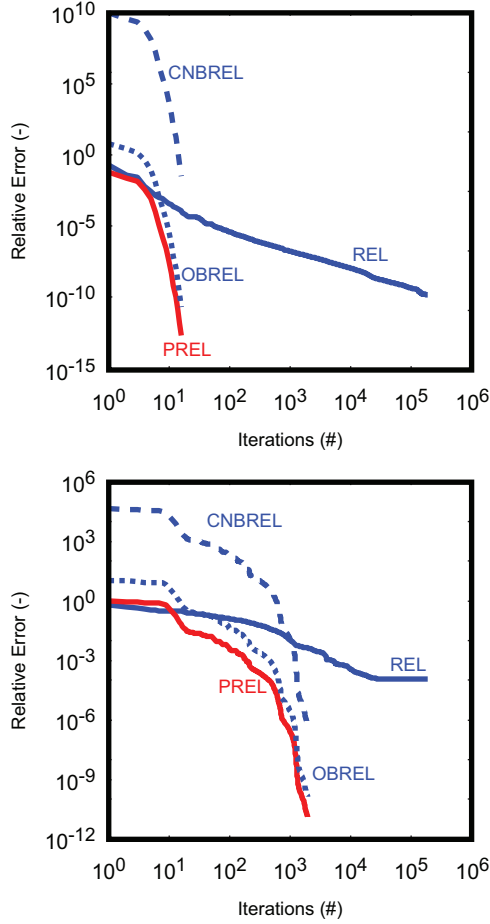
### Model Trust Region Methods

Let  $\mathbf{F} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ . Assume that for steps  $\mathbf{s}$  satisfying  $\|\mathbf{s}\|_2 \leq \delta$ , the function

$$\hat{m}_{\mathbf{x}}(\mathbf{s}) = \left(\frac{1}{2}\right) \|\mathbf{F}(\mathbf{x})\|_2^2 + ((D\mathbf{F})(\mathbf{x})^\top \mathbf{F}(\mathbf{x}))^\top \mathbf{s} + \left(\frac{1}{2}\right) \mathbf{s}^\top (D\mathbf{F})(\mathbf{x})^\top (D\mathbf{F})(\mathbf{x}) \mathbf{s}$$

is a suitable *local model* of the globalizing objective function  $f(\mathbf{x}) = \|\mathbf{F}(\mathbf{x})\|_2^2/2$ .<sup>5</sup> The

<sup>5</sup>Note that this is not the usual, quadratic model of  $f$ , because  $(D\mathbf{F})(\mathbf{x})^\top (D\mathbf{F})(\mathbf{x}) \neq (D\nabla f)(\mathbf{x})$



**Figure C.1** Relative error in computed solutions to the Newton system and its preconditioned form using GMRES under the [Berry et al. \(1995\)](#) model. On the top, prices are  $\mathbf{p} = \mathbf{p}^* + 100\boldsymbol{\nu}$  where  $\mathbf{p}^*$  are equilibrium prices and  $\boldsymbol{\nu} \in [-\mathbf{1}, \mathbf{1}]$  is a sample from a uniformly distributed random vector. For this case  $\kappa(\boldsymbol{\Lambda}(\mathbf{p})) = 1.56 \times 10^{11}$  while the multiplier in Eqn. (C.8) is only 106.41. On the bottom, prices are  $\mathbf{p} = 20,000\boldsymbol{\nu} + 5,000$  where  $\boldsymbol{\nu}$  is a sample from a random vector uniformly distributed on  $[\mathbf{0}, \mathbf{1}]$ . For this case  $\kappa(\boldsymbol{\Lambda}(\mathbf{p})) = 4.6 \times 10^4$  while the multiplier in Eqn. (C.8) is only 10.73. Abbreviations are as follows. REL: relative error in the Newton System; PREL: relative error in the pre-conditioned Newton System; OBREL: our bound, Eqn. (C.8), on the relative error in the Newton System as determined from the relative error in the preconditioned Newton system; CNBREL: condition number bound on the relative error in the Newton System as determined from the relative error in the preconditioned Newton system.

idea is to solve

$$\min_{\|\mathbf{s}\|_2 \leq \delta} \hat{m}_{\mathbf{x}}(\mathbf{s}). \quad (\text{C.9})$$

The solution  $\mathbf{s}_*$  is given as follows: take  $\mathbf{s}_* = \mathbf{s}^N = -(DF)(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})$  if  $\|\mathbf{s}^N\|_2 \leq \delta$ ; if  $\|\mathbf{s}^N\|_2 > \delta$ , take  $\mathbf{s}_* = \mathbf{s}(\mu_*)$  where

$$\mathbf{s}(\mu) = -((DF)(\mathbf{x})^\top (DF)(\mathbf{x}) + \mu \mathbf{I})^{-1} (DF)(\mathbf{x})^\top \mathbf{F}(\mathbf{x})$$

and  $\mu_* > 0$  is the unique  $\mu > 0$  such that  $\|\mathbf{s}(\mu)\|_2 = \delta$ .<sup>6</sup>

### Hybrid Model Trust Region - Krylov Subspace Methods

A Krylov method for solving  $(DF)(\mathbf{x})\mathbf{s}^N = -\mathbf{F}(\mathbf{x})$  builds approximate solutions in the successive *Krylov subspaces*  $\mathcal{K}^{(n)} = \text{span}\{-(DF)(\mathbf{x})^m \mathbf{F}(\mathbf{x})\}_{m=0}^{n-1}$  (Trefethen and Bau, 1997). This has the effect of further constraining the local model problem (C.9) to

$$\min_{\mathbf{s} \in \mathcal{K}^{(n)}, \|\mathbf{s}\|_2 \leq \delta} \hat{m}_{\mathbf{x}}(\mathbf{s}). \quad (\text{C.10})$$

As we describe in greater detail below GMRES builds an orthonormal basis for  $\mathcal{K}^{(n)}$ , contained in the columns of  $\mathbf{Q}^{(n)} \in \mathbb{R}^{J \times n}$ . For any  $\mathbf{Q} \in \mathbb{R}^{J \times n}$  with orthonormal columns (generated by GMRES or not) we can set  $\hat{m}_{\mathbf{x}, \mathbf{Q}}(\mathbf{q}) = \hat{m}_{\mathbf{x}}(\mathbf{Q}\mathbf{q})$  and restrict attention to the trust region problem  $\min_{\|\mathbf{q}\|_2 \leq \delta} \hat{m}_{\mathbf{x}, \mathbf{Q}}(\mathbf{q})$ .<sup>7</sup> The first-order conditions for this problem<sup>8</sup> are equivalent to either

- (i)  $(\nabla \hat{m}_{\mathbf{x}, \mathbf{Q}})(\mathbf{q}) = \mathbf{0}$  and  $\|\mathbf{q}\|_2 \leq \delta$
- (ii) or  $(\nabla \hat{m}_{\mathbf{x}, \mathbf{Q}})(\mathbf{q}) + \mu \mathbf{q} = \mathbf{0}$  for  $\|\mathbf{q}\|_2 = \delta$  and some  $\mu > 0$ .

By the definition of  $\hat{m}_{\mathbf{x}, \mathbf{Q}}$ , (i) implies

$$\mathbf{Q}^\top (DF)(\mathbf{x})^\top (DF)(\mathbf{x}) \mathbf{Q} \mathbf{q} + \mathbf{Q}^\top (DF)(\mathbf{x})^\top \mathbf{F}(\mathbf{x}) = \mathbf{0}$$

and (ii) implies

$$\mathbf{Q}^\top (DF)(\mathbf{x})^\top (DF)(\mathbf{x}) \mathbf{Q} \mathbf{q} + \mu \mathbf{q} + \mathbf{Q}^\top (DF)(\mathbf{x})^\top \mathbf{F}(\mathbf{x}) = \mathbf{0}.$$

Note that these are square problems that can be solved exactly.

---

(Dennis and Schnabel, 1996, pg. 149).

<sup>6</sup>These follow from the standard optimality conditions, or rather that the gradient  $(\nabla \hat{m}_{\mathbf{x}})(\mathbf{s})$  must lie in the negative normal cone to  $\bar{\mathbb{B}}_\delta(\mathbf{0}) = \{\mathbf{y} \in \mathbb{R}^N : \|\mathbf{y}\|_2 \leq \delta\}$  at  $\mathbf{x}$  (Clarke, 1975); see (Dennis and Schnabel, 1996, Lemma 6.4.1, pg. 131).

<sup>7</sup>See (Brown and Saad, 1990, pgs. 149-150).

<sup>8</sup>Again, the gradient of  $\hat{m}_{\mathbf{x}, \mathbf{Q}}$  lies in the negative normal cone to  $\bar{\mathbb{B}}_\delta(\mathbf{0}) = \{\mathbf{y} \in \mathbb{R}^N : \|\mathbf{y}\|_2 \leq \delta\}$  at  $\mathbf{x}$  (Clarke, 1975; Dennis and Schnabel, 1996).

## Householder GMRES

As described by [Saad and Schultz \(1986\)](#), [Golub and Loan \(1996\)](#), and [Trefethen and Bau \(1997\)](#), the GMRES process is based on the classical Gram-Schmidt orthogonalization process and unstable. We implement a numerically stable variant based on Householder transformations due to [Walker \(1988\)](#).<sup>9</sup> In this version of the GMRES process applied to the generic problem  $\mathbf{Ax} = \mathbf{b}$ , Householder reflectors  $\mathbf{P}^{(n)} \in \mathbb{R}^{N \times N}$  are used to generate the orthonormal matrices

$$\mathbf{Q}^{(n)} = \mathbf{P}^{(1)} \dots \mathbf{P}^{(n)} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{N \times n} \quad (\mathbf{I} \in \mathbb{R}^{n \times n}, \mathbf{0} \in \mathbb{R}^{(N-n) \times n})$$

satisfying

$$\mathbf{A}\mathbf{Q}^{(n)} = \mathbf{P}^{(1)} \dots \mathbf{P}^{(n+1)} \mathbf{H}^{(n)} = \mathbf{Q}^{(n+1)} \tilde{\mathbf{H}}^{(n)}$$

where  $\mathbf{H}^{(n)} \in \mathbb{R}^{N \times n}$  is

$$\mathbf{H}^{(n)} = \begin{bmatrix} \tilde{\mathbf{H}}^{(n)} \\ \mathbf{0} \end{bmatrix}$$

for upper Hessenberg  $\tilde{\mathbf{H}}^{(n)} \in \mathbb{R}^{(n+1) \times n}$  and  $\mathbf{0} \in \mathbb{R}^{(N-n-1) \times n}$ .  $\mathbf{P}^{(1)}$  is chosen to satisfy  $\mathbf{P}^{(1)}\mathbf{b} = -\beta\mathbf{e}_1$  where  $\beta = \text{sign}(b_1)\|\mathbf{b}\|_2$ , and hence  $(\mathbf{Q}^{(n)})^\top \mathbf{b} = -\beta\mathbf{e}_1$ . The  $n^{\text{th}}$  approximate solution  $\mathbf{x}^{(n)}$  is taken to be  $\mathbf{x}^{(n)} = \mathbf{Q}^{(n)}\mathbf{y}^{(n)}$  where  $\mathbf{y}^{(n)} \in \mathbb{R}^n$  solves

$$\min_{\mathbf{y} \in \mathbb{R}^n} \|\tilde{\mathbf{H}}^{(n)}\mathbf{y} - \beta\mathbf{e}_1\|_2.$$

These problems can be solved cheaply by updating Givens QR factorizations, and the solution vector and residual need not be formed until GMRES converges.

For  $\mathbf{A} = (D\mathbf{F})(\mathbf{x})$  and  $\mathbf{b} = -\mathbf{F}(\mathbf{x})$ ,  $\beta = -\text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2$  and  $-\beta\mathbf{e}_1 = \mathbf{P}^{(1)}\mathbf{b} = -\mathbf{P}^{(1)}\mathbf{F}(\mathbf{x})$  so that

$$\mathbf{P}^{(1)}\mathbf{F}(\mathbf{x}) = \beta\mathbf{e}_1 = -\text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2\mathbf{e}_1.$$

Moreover,  $\mathbf{P}^{(n)}\mathbf{e}_1 = \mathbf{e}_1$  for all  $n > 1$  so that

$$(\mathbf{Q}^{(n)})^\top \mathbf{F}(\mathbf{x}) = -\text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2\mathbf{e}_1.$$

---

<sup>9</sup>This is also the version implemented in `matlab`. We have verified that our implementation generates results matching `matlab`'s implementation.

## Hybrid Model Trust Region - GMRES Methods

Using GMRES started at zero,  $(DF)(\mathbf{x})\mathbf{Q}^{(n)} = \mathbf{Q}^{(n+1)}\tilde{\mathbf{H}}^{(n)}$  and  $(\mathbf{Q}^{(n+1)})^\top \mathbf{F}(\mathbf{x}) = -\text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 \mathbf{e}_1$ . Thus we consider the family of  $n \times n$  linear systems

$$\begin{aligned} & (\mathbf{Q}^{(n)})^\top (DF)(\mathbf{x})^\top (DF)(\mathbf{x})\mathbf{Q}^{(n)} \mathbf{q} + \mu \mathbf{q} + (\mathbf{Q}^{(n)})^\top (DF)(\mathbf{x})^\top \mathbf{F}(\mathbf{x}) \\ & = ((\tilde{\mathbf{H}}^{(n)})^\top \tilde{\mathbf{H}}^{(n)} + \mu \mathbf{I}) \mathbf{q} - \text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 (\tilde{\mathbf{H}}^{(n)})^\top \mathbf{e}_1 = \mathbf{0} \end{aligned}$$

defined for all  $\mu \geq 0$ .

By computing the (“thin”) Singular Value Decomposition of  $\tilde{\mathbf{H}}^{(n)}$ ,  $\tilde{\mathbf{H}}^{(n)} = \tilde{\mathbf{U}}\Sigma\mathbf{V}^\top$  where  $\tilde{\mathbf{U}} \in \mathbb{R}^{(n+1) \times n}$ ,  $\mathbf{V} \in \mathbb{R}^{n \times n}$ , and  $\Sigma \in \mathbb{R}^{n \times n}$ , we can easily solve each such problem.<sup>10</sup> Particularly,

$$((\tilde{\mathbf{H}}^{(n)})^\top \tilde{\mathbf{H}}^{(n)} + \mu \mathbf{I}) \mathbf{q} - \text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 (\tilde{\mathbf{H}}^{(n)})^\top \mathbf{e}_1 = \mathbf{0}$$

is solved by  $\mathbf{q}(\mu) = \mathbf{V}\boldsymbol{\eta}(\mu)$  where

$$\boldsymbol{\eta}(\mu) = \text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 (\Sigma^2 + \mu \mathbf{I})^{-1} \Sigma \tilde{\mathbf{U}}^\top \mathbf{e}_1.$$

Because the diagonal elements of  $\Sigma^2$  are positive,  $\boldsymbol{\eta}(\mu)$  is well defined for all  $\mu \geq 0$ . Note also that we only need the first row of  $\mathbf{U}$ , but all of  $\mathbf{V}$ , to compute  $\mathbf{q}(\mu)$ .

In particular,  $\mathbf{q}(0) = \text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 \mathbf{V}\Sigma^{-1}\mathbf{U}^\top \mathbf{e}_1$ . Invoking the *full* SVD of  $\tilde{\mathbf{H}}^{(n)}$ ,

$$\tilde{\mathbf{H}}^{(n)} = \begin{bmatrix} \tilde{\mathbf{U}} & \mathbf{u}_{n+1} \end{bmatrix} \begin{bmatrix} \Sigma \\ \mathbf{0}^\top \end{bmatrix} \mathbf{V}^\top$$

for some  $\mathbf{u}_{n+1} \perp \text{span}\{\mathbf{u}_i\}_{i=1}^n$ , we can write

$$\|\tilde{\mathbf{H}}^{(n)} \mathbf{q} - \text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 \mathbf{e}_1\|_2 = \left\| \begin{bmatrix} \tilde{\Sigma} \mathbf{V}^\top \mathbf{q} \\ 0 \end{bmatrix} - \text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 \begin{bmatrix} \tilde{\mathbf{U}}^\top \mathbf{e}_1 \\ u_{1,n+1} \end{bmatrix} \right\|_2.$$

We thus see that  $\mathbf{q}(0)$  solves the  $(n+1) \times n$  GMRES least squares problem

$$\min_{\mathbf{q}} \|\mathbf{H}^{(n+1,n)} \mathbf{q} - \text{sign}(F_1(\mathbf{x}))\|\mathbf{F}(\mathbf{x})\|_2 \mathbf{e}_1\|_2.$$

with residual  $|u_{1,n+1}| \|\mathbf{F}(\mathbf{x})\|_2$ .<sup>11</sup>

<sup>10</sup>See (Golub and Loan, 1996, Section 12.1, pgs. 580-583) for closely related results.

<sup>11</sup> $|u_{1,n+1}|$  is unique: First, note that  $\mathbf{u}_{n+1}$  is a unit vector in the span of a single vector, say  $\mathbf{v}$ ,

It is also easy to see that

$$\begin{aligned}
\mathbf{F}(\mathbf{x})^\top (D\mathbf{F})(\mathbf{x})\mathbf{s}^{(n)}(\mu) &= \mathbf{F}(\mathbf{x})^\top (D\mathbf{F})(\mathbf{x})\mathbf{Q}^{(n)}\mathbf{q}^{(n)}(\mu) \\
&= \left( (\mathbf{Q}^{(n+1)})^\top \mathbf{F}(\mathbf{x}) \right)^\top \tilde{\mathbf{H}}^{(n)}\mathbf{q}^{(n)}(\mu) \\
&= -\beta^2 (\boldsymbol{\nu}_1^\top \mathbf{D}(\mu)\boldsymbol{\nu}_1) \\
&= -\|\mathbf{F}(\mathbf{x})\|_2^2 (\boldsymbol{\nu}_1^\top \mathbf{D}(\mu)\boldsymbol{\nu}_1) < 0
\end{aligned}$$

where  $\boldsymbol{\nu}_1$  is the first row of  $\tilde{\mathbf{U}}$  and  $\mathbf{D}(\mu) = \text{diag}(d_1(\mu), \dots, d_n(\mu))$  for  $d_i(\mu) = \sigma_i^2/(\sigma_i^2 + \mu)$ . That is, *the Householder GMRES Newton-Hookstep always lies in a descent direction for the globalizing objective  $f(\mathbf{x}) = \|\mathbf{F}(\mathbf{x})\|_2^2/2$ .*

### Trust Region Constrained Steps

It remains to find  $\mu_* > 0$  such that  $\mathbf{q}_* = \mathbf{q}(\mu_*)$  satisfies  $\|\mathbf{q}_*\|_2 = \delta$  when  $\|\mathbf{q}(0)\|_2 > \delta$ . First note that  $\|\mathbf{q}(\mu)\|_2 = \delta$  if and only if  $\|\boldsymbol{\eta}(\mu)\|_2 = \delta$ . Next we define

$$\begin{aligned}
\psi(\mu) &= \left( \frac{1}{2} \right) \left[ \|\boldsymbol{\eta}(\mu)\|_2^2 - \left( \frac{\delta}{\|\mathbf{F}(\mathbf{x})\|_2} \right)^2 \right] \\
&= \left( \frac{1}{2} \right) \left[ \sum_{i=1}^n \left( \frac{\sigma_i u_{1,i}}{\sigma_i^2 + \mu} \right)^2 - \left( \frac{\delta}{\|\mathbf{F}(\mathbf{x})\|_2} \right)^2 \right].
\end{aligned}$$

The proofs of the following easy claims are left to the reader.

**Lemma C.2.9.** (i)  $\|\mathbf{q}(0)\|_2 = \|\boldsymbol{\eta}(0)\|_2 > \delta$  if and only if  $\psi(0) > 0$ . (ii)  $\lim_{\mu \rightarrow \infty} \psi(\mu) < 0$ . (iii)  $(D\psi)(\mu) < 0$  for all  $\mu \geq 0$ . (iv)  $(D^2\psi)(\mu) > 0$  for all  $\mu \geq 0$ .

The differentiation denoted by “ $D$ ” is, of course, taken with respect to  $\mu$ . These results establish the following uniqueness result.

**Corollary C.2.10.** *If  $\psi(0) > 0$ , then  $\psi$  has a unique root  $\mu_* > 0$ .*

There are two obvious ways to solve for  $\mu_* > 0$  when  $\psi(0) > 0$ . The first is Newton’s method, a linear local model approach, and the second is a more general nonlinear local model approach inspired by Newton’s method and the specific form of  $\psi$ . See (Dennis and Schnabel, 1996, pgs. 134-136).

---

that is orthogonal to the span of the columns of  $\tilde{\mathbf{U}}$ . There are only two unit vectors in this span, specifically  $\pm \mathbf{v}/\|\mathbf{v}\|_2$ , and thus  $\mathbf{u}_{n+1} \in \{\pm \mathbf{v}/\|\mathbf{v}\|_2\}$ . Thus  $|u_{1,n+1}| \in \{|\pm v_1/\|\mathbf{v}\|_2| = |v_1|/\|\mathbf{v}\|_2\}$ .

Newton's method is easily seen to be update  $\mu \leftarrow \mu + \nu(\mu)$  where

$$\nu(\mu) = -\frac{\psi(\mu)}{(D\psi)(\mu)} = \left(\frac{1}{2}\right) \left(\frac{\sum_{i=1}^n \sigma_i^2 u_{1,i}^2 / (\sigma_i^2 + \mu)^2 - (\delta / \|\mathbf{F}(\mathbf{x})\|_2)^2}{\sum_{i=1}^n \sigma_i^2 u_{1,i}^2 / (\sigma_i^2 + \mu)^3}\right).$$

Clearly  $\|\boldsymbol{\eta}(\mu)\|_2 > \delta / \|\mathbf{F}(\mathbf{x})\|_2$  implies  $\nu(\mu) > 0$  and  $\|\boldsymbol{\eta}(\mu)\|_2 < \delta / \|\mathbf{F}(\mathbf{x})\|_2$  implies  $\nu(\mu) < 0$ . In other words,  $\mu$  is increased (decreasing  $\|\boldsymbol{\eta}(\mu)\|_2$ ) whenever  $\|\boldsymbol{\eta}(\mu)\|_2 > \delta / \|\mathbf{F}(\mathbf{x})\|_2$  and  $\mu$  is decreased (increasing  $\|\boldsymbol{\eta}(\mu)\|_2$ ) whenever  $\|\boldsymbol{\eta}(\mu)\|_2 < \delta / \|\mathbf{F}(\mathbf{x})\|_2$ .

For the nonlinear local model approach, we define the local model

$$\hat{\psi}_\mu(\hat{\mu}) = \left(\frac{\alpha}{\beta + \mu}\right)^2 - \left(\frac{\delta}{\|\mathbf{F}(\mathbf{x})\|_2}\right)^2$$

for some  $\alpha, \beta$ . To choose  $\alpha, \beta$ , we fit  $\hat{\psi}_\mu$  to  $\psi$  by requiring (a)  $\hat{\psi}_\mu(\mu) = \psi(\mu)$  and (b)  $(D\hat{\psi}_\mu)(\mu) = (D\psi)(\mu)$ . In other words,  $\alpha / (\beta + \mu) = \|\boldsymbol{\eta}(\mu)\|_2$  and

$$\frac{\alpha^2}{(\beta + \mu)^3} = \sum_{i=1}^n \frac{\sigma_i^2 u_{1,i}^2}{(\sigma_i^2 + \mu)^3}.$$

The first condition implies  $\alpha / \|\boldsymbol{\eta}(\mu)\|_2 = \beta + \mu$ , and the second condition becomes

$$\frac{\alpha}{\|\boldsymbol{\eta}(\mu)\|_2} = \frac{\|\boldsymbol{\eta}(\mu)\|_2^2}{\sum_{i=1}^n \sigma_i^2 u_{1,i}^2 / (\sigma_i^2 + \mu)^3}.$$

As Newton's method chooses an update to zero the linear local model, here we choose an update to zero the nonlinear local model  $\hat{\psi}_\mu$ . Note that  $\hat{\psi}_\mu(\hat{\mu}_0) = 0$  for

$$\hat{\mu}_0 = \left(\frac{\|\mathbf{F}(\mathbf{x})\|_2}{\delta}\right) \alpha - \beta.$$

Thus we generate the update  $\mu \leftarrow \mu + \nu(\mu)$  where

$$\nu(\mu) = \left(\frac{\|\boldsymbol{\eta}(\mu)\|_2^2}{\sum_{i=1}^n \sigma_i^2 u_{1,i}^2 / (\sigma_i^2 + \mu)^3}\right) \left(\frac{\|\boldsymbol{\eta}(\mu)\|_2}{\delta / \|\mathbf{F}(\mathbf{x})\|_2} - 1\right).$$

Again, if  $\|\boldsymbol{\eta}(\mu)\|_2 > \delta / \|\mathbf{F}(\mathbf{x})\|_2$  then  $\mu$  is increased (decreasing  $\|\boldsymbol{\eta}(\mu)\|_2$ ), while if  $\|\boldsymbol{\eta}(\mu)\|_2 < \delta / \|\mathbf{F}(\mathbf{x})\|_2$  then  $\mu$  is decreased (increasing  $\|\boldsymbol{\eta}(\mu)\|_2$ ).

## Preconditioning

An important component of successful implementations of GMRES is preconditioning. Let  $\mathbf{M} = \mathbf{M}_L \mathbf{M}_R$  be a preconditioner. The local quadratic model (prior to projec-

tion onto a Krylov subspace) was half of the squared 2-norm of the linear model  $\ell_{\mathbf{x}}(\mathbf{s}) = \mathbf{F}(\mathbf{x}) + (D\mathbf{F})(\mathbf{x})\mathbf{s}$ . Since this locally linear model has a zero if and only if the preconditioned local linear model

$$\ell_{\mathbf{x},\mathbf{M}}(\mathbf{s}) = \mathbf{M}^{-1}\mathbf{F}(\mathbf{x}) + \mathbf{M}^{-1}(D\mathbf{F})(\mathbf{x})\mathbf{s}$$

does, we can consider the preconditioned local quadratic model

$$\begin{aligned} \hat{m}_{\mathbf{x},\mathbf{M}}(\mathbf{s}) &= \left(\frac{1}{2}\right) \|\mathbf{M}^{-1}\mathbf{F}(\mathbf{x})\|_2^2 + \mathbf{F}(\mathbf{x})^\top \mathbf{M}^{-\top} \mathbf{M}^{-1} (D\mathbf{F})(\mathbf{x})\mathbf{s} \\ &\quad + \left(\frac{1}{2}\right) \mathbf{s}^\top (D\mathbf{F})(\mathbf{x})^\top \mathbf{M}^{-\top} \mathbf{M}^{-1} (D\mathbf{F})(\mathbf{x})\mathbf{s} \end{aligned}$$

of the globalizing objective function  $f_{\mathbf{M}}(\mathbf{x}) = \|\mathbf{M}^{-1}\mathbf{F}(\mathbf{x})\|_2^2/2$ .

Projecting this model into a subspace spanned by the columns of any matrix  $\mathbf{Q}$ ,

$$\begin{aligned} \hat{m}_{\mathbf{x},\mathbf{M}}^{\mathbf{Q}}(\mathbf{q}) &= \left(\frac{1}{2}\right) \|\mathbf{M}^{-1}\mathbf{F}(\mathbf{x})\|_2^2 + \mathbf{F}(\mathbf{x})^\top \mathbf{M}^{-\top} \mathbf{M}^{-1} (D\mathbf{F})(\mathbf{x})\mathbf{Q}\mathbf{q} \\ &\quad + \left(\frac{1}{2}\right) \mathbf{q}^\top \mathbf{Q}^\top (D\mathbf{F})(\mathbf{x})^\top \mathbf{M}^{-\top} \mathbf{M}^{-1} (D\mathbf{F})(\mathbf{x})\mathbf{Q}\mathbf{q}. \end{aligned}$$

Defining  $\mathbf{q}(\mu)$  by

$$(\mathbf{Q}^\top (D\mathbf{F})(\mathbf{x})^\top \mathbf{M}^{-\top} \mathbf{M}^{-1} (D\mathbf{F})(\mathbf{x})\mathbf{Q} + \mu\mathbf{I})\mathbf{q}(\mu) + \mathbf{Q}^\top (D\mathbf{F})(\mathbf{x})^\top \mathbf{M}^{-\top} \mathbf{M}^{-1}\mathbf{F}(\mathbf{x}) = \mathbf{0}$$

the constrained optimality conditions again state that the solution to the projected trust region problem is given by  $\mathbf{q}(0)$  if  $\|\mathbf{q}(0)\|_2 \leq \delta$  and otherwise by  $\mathbf{q}(\mu_*)$  for the unique  $\mu_*$  such that  $\|\mathbf{q}(\mu)\|_2 = \delta$ .

Applying GMRES to the preconditioned system  $\mathbf{M}^{-1}(D\mathbf{F})(\mathbf{x})\mathbf{s} = -\mathbf{M}^{-1}\mathbf{F}(\mathbf{x})$ , we build orthonormal matrices  $\mathbf{Q}_{\mathbf{M}}^{(n)}$  satisfying  $\mathbf{M}^{-1}(D\mathbf{F})(\mathbf{x})\mathbf{Q}_{\mathbf{M}}^{(n)} = \mathbf{Q}_{\mathbf{M}}^{(n+1)}\mathbf{H}_{\mathbf{M}}^{(n)}$  and  $(\mathbf{Q}_{\mathbf{M}}^{(n)})^\top \mathbf{M}^{-1}\mathbf{F}(\mathbf{x}) = \text{sign}((\mathbf{M}^{-1}\mathbf{F}(\mathbf{x}))_1) \|\mathbf{M}^{-1}\mathbf{F}(\mathbf{x})\|_2 \mathbf{e}_1$  for all  $n$ . Thus  $\mathbf{q}^{(n)}(\mu)$  satisfies

$$((\mathbf{H}_{\mathbf{M}}^{(n)})^\top \mathbf{H}_{\mathbf{M}}^{(n)} + \mu\mathbf{I})\mathbf{q}^{(n)}(\mu) - \text{sign}((\mathbf{M}^{-1}\mathbf{F}(\mathbf{x}))_1) \|\mathbf{M}^{-1}\mathbf{F}(\mathbf{x})\|_2 (\mathbf{H}_{\mathbf{M}}^{(n)})^\top \mathbf{e}_1 = \mathbf{0}.$$

Invoking the (thin) SVD of  $\mathbf{H}_{\mathbf{M}}^{(n)}$ ,  $\mathbf{H}_{\mathbf{M}}^{(n)} = \mathbf{U}_{\mathbf{M}}\mathbf{\Sigma}_{\mathbf{M}}\mathbf{V}_{\mathbf{m}}^\top$ ,

$$\mathbf{q}^{(n)}(\mu) = \text{sign}((\mathbf{M}^{-1}\mathbf{F}(\mathbf{x}))_1) \|\mathbf{M}^{-1}\mathbf{F}(\mathbf{x})\|_2 \mathbf{V}_{\mathbf{M}}(\mathbf{\Sigma}_{\mathbf{M}}^2 + \mu\mathbf{I})^{-1} \mathbf{\Sigma}_{\mathbf{M}} \mathbf{U}_{\mathbf{M}}^\top \mathbf{e}_1.$$

Only one potential problem arises when using preconditioning. While the hooksteps obtained lie in descent directions for the preconditioned globalizing objective, they do



not necessarily lie in descent directions for the original globalizing objective.

### A Limited-Memory BFGS-Inexact Newton Method

There is a potential variation of the BFGS Quasi-Newton method that can be applied. Let  $\mathbf{A}^{(0)}$  be an initial approximation to the Jacobian  $(D\mathbf{F})(\mathbf{x}^{(0)})$ . In each iteration  $n \geq 1$ , we have  $\mathbf{F}^{(n)} = \mathbf{F}(\mathbf{x}^{(n)})$ ,  $\mathbf{s}^{(n)} = \mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}$  (which may be different from the Newton step because of the globalization strategy),  $\hat{\mathbf{s}}^{(n)} = \mathbf{s}^{(n)} / \|\mathbf{s}^{(n)}\|_2^2$  and  $\mathbf{d}^{(n)} = \mathbf{F}^{(n)} - (\mathbf{F}^{(n-1)} + \mathbf{A}^{(n-1)}\mathbf{s}^{(n)})$ . Products  $\mathbf{A}^{(n)}\mathbf{y}$  can be calculated using the formula

$$\begin{aligned} \mathbf{A}^{(n)}\mathbf{y} &= \mathbf{A}^{(n-1)}\mathbf{y} + (\mathbf{y}^\top \hat{\mathbf{s}}^{(n)}) \mathbf{d}^{(n)} \\ &= \mathbf{A}^{(n-2)}\mathbf{y} + (\mathbf{y}^\top \hat{\mathbf{s}}^{(n-1)}) \mathbf{d}^{(n-1)} + (\mathbf{y}^\top \hat{\mathbf{s}}^{(n)}) \mathbf{d}^{(n)} \\ &\quad \vdots \\ &= \mathbf{A}^{(0)}\mathbf{y} + \sum_{m=1}^n (\mathbf{y}^\top \hat{\mathbf{s}}^{(m)}) \mathbf{d}^{(m)} \\ &= \mathbf{A}^{(0)}\mathbf{y} + \mathbf{D}^{(n)}(\mathbf{S}^{(n)})^\top \mathbf{y} \end{aligned}$$

from  $\mathbf{A}^{(0)}$  and  $\{(\hat{\mathbf{s}}^{(m)}, \mathbf{d}^{(m)})\}_{m=1}^n$ , where  $\mathbf{S}^{(n)} \in \mathbb{R}^{n \times J}$  has columns  $\{\hat{\mathbf{s}}^{(m)}\}_{m=1}^n$ , and  $\mathbf{D}^{(n)} \in \mathbb{R}^{J \times n}$  has columns  $\{\mathbf{d}^{(m)}\}_{m=1}^n$ .<sup>12</sup> These products can be used by GMRES to solve for a BFGS-inexact Newton step  $\mathbf{s}^{IN}$  satisfying  $\|\mathbf{F}^{(n)} + \mathbf{A}^{(n)}\mathbf{s}^{IN}\|_2 \leq \delta \|\mathbf{F}^{(n)}\|_2$  for some  $\delta > 0$ . Viswanath's GMRES-Hookstep can then be applied, resulting in the (acceptable) step  $\mathbf{s}^{(n+1)}$ , new prices  $\mathbf{p}^{(n+1)} = \mathbf{p}^{(n)} + \mathbf{s}^{(n+1)}$ , and new function value  $\mathbf{F}^{(n+1)} = \mathbf{F}(\mathbf{p}^{(n+1)})$ . Practically, a limited number of the vector pairs  $(\hat{\mathbf{s}}^{(m)}, \mathbf{d}^{(m)})$  should be stored.

We have given an outline of a “limited-memory” BFGS-inexact Newton method (Nocedal and Liu, 1989) with GMRES-Hookstep. If the truncated BFGS approximations to the Jacobian are good enough, this approach could result in a significant reduction in computational burden by eliminating the need to compute  $(D\mathbf{F})(\mathbf{x}^{(n)})$  or evaluate  $\mathbf{F}$  in directional finite differences during the GMRES process. However, computational experience with the BFGS Quasi-Newton method applied to solve  $(\tilde{\nabla} \hat{\pi})(\mathbf{p}) = \mathbf{0}$  leads us to suspect that the rate of convergence will be dramatically slowed, potentially usurping these reductions in computational burden.

---

<sup>12</sup>The order in which the vectors  $\mathbf{s}^{(m)}$  and  $\mathbf{d}^{(m)}$  appear as columns in  $\mathbf{S}^{(n)}$  and  $\mathbf{D}^{(n)}$  is irrelevant, so long as  $(\mathbf{S}^{(n)})_{\cdot, l} = \hat{\mathbf{s}}^{(m)}$  if, and only if,  $(\mathbf{D}^{(n)})_{\cdot, l} = \mathbf{d}^{(m)}$ .

# Appendix D

## Proofs

### D.1 Proofs for Chapter 3

#### Existence

*Proof of Proposition 3.1.2.* The formula Eqn. (3.2) follows from the chain rule in the usual way; this is a generalization of the formula in (Train, 2003, Chapter 3). (i), (ii), and (iii) follow from the choice probability formula Eqn. (3.1). For (iv) we use Lemma 3.1.3.

We must also show that  $\mathbf{P} : [\mathbf{0}, \infty) \rightarrow \Delta(J)$  is onto to prove (v). Let  $\mathbf{x} \in \Delta(J)$ . Since

$$\frac{e^{u_j(p_j)}}{\sum_{k=1}^J e^{u_k(p_k)}} = \frac{e^\lambda e^{u_j(p_j)}}{\sum_{k=1}^J e^\lambda e^{u_k(p_k)}} = \frac{e^{u_j(p_j)+\lambda}}{\sum_{k=1}^J e^{u_k(p_k)+\lambda}}$$

for any  $\lambda \in \mathbb{R}$  it suffices to solve

$$\begin{aligned} u_1(p_1) &= \log x_1 - \lambda \\ &\vdots \\ u_J(p_J) &= \log x_J - \lambda \end{aligned}$$

for some  $\lambda \in \mathbb{R}$ . A necessary condition is that  $\log x_j - \lambda \leq u_j(0)$  for all  $j$ . Particularly, if we choose  $\lambda \geq -\min_k u_k(0)$  then  $\log x_j - \lambda \leq \min_k u_k(0) \leq u_j(0)$  for all  $j$ . Because every  $u_j$  is strictly decreasing, there is then a unique  $\mathbf{p}$  that satisfies the nonlinear system of equations above.  $\square$

*Proof of Lemma 3.1.3.* By definition,  $w(\chi_{j,p_j}(\lambda)) = w_j(p_j) - \log \lambda$ . Now if  $\lambda' > \lambda$ , then

$$w(\chi_{j,p_j}(\lambda)) = w_j(p_j) - \log \lambda > w_j(p_j) - \log \lambda' = w(\chi_{j,p_j}(\lambda'))$$

Because  $w_j$  is strictly decreasing,  $\chi_{j,p_j}(\lambda') > \chi_{j,p_j}(\lambda)$ . This proves the first part of (i). The second follows from the fact that

$$\lim_{\lambda \rightarrow \infty} w(\chi_{j,p_j}(\lambda)) = w_j(p_j) - \lim_{\lambda \rightarrow \infty} \log \lambda = -\infty.$$

Since  $e^{u(\chi_{j,p_j}(\lambda))} = \lambda e^{w_j(p_j)}$  (ii) and (iii) follow immediately from the Logit choice probability formulas.  $\square$

*Proof of Proposition 3.1.4.* The componentwise formula Eqn. (3.2) is straightforward to evaluate from Eqn. (3.1), and the matrix forms follow easily. See Anderson and de Palma (1992a) or Train (2003) for similar formula for specific utility functions. Eqn. (3.4) is a bit tedious to derive from Eqn. (3.2), but not any more difficult.  $\square$

*Proof of Proposition 3.1.5.* The key is to observe that

$$\lim_{p_j \rightarrow q_j} p_j e^{u_j(p_j) - \vartheta} < \infty \quad \text{implies} \quad \lim_{\mathbf{p} \rightarrow \mathbf{q}} p_j P_j^L(\mathbf{p}) < \infty,$$

or more specifically,

$$\lim_{p_j \rightarrow q_j} p_j e^{u_j(p_j) - \vartheta} = 0 \quad \text{implies} \quad \lim_{\mathbf{p} \rightarrow \mathbf{q}} p_j P_j^L(\mathbf{p}) = 0.$$

For

$$p_j P_j^L(\mathbf{p}) = \frac{p_j e^{u_j(p_j)}}{e^\vartheta + \sum_{k=1}^J e^{u_k(p_k)}} \leq p_j e^{u_j(p_j) - \vartheta}.$$

Under the first bounding hypothesis,  $e^{u_j(p_j)} \leq p^{-r_j} e^{\kappa_j + v_j}$  for  $p_j > \bar{p}_j$  where  $\kappa_j = \kappa(\mathbf{y}_j)$ ,  $r_j = r(\mathbf{y}_j)$ , and  $\bar{p}_j = \bar{p}(\mathbf{y}_j)$ . Thus  $p_j e^{u_j(p_j)} \leq p^{1-r_j} e^{\kappa_j + v_j}$ . If  $r_j \geq 1$ , this tends to a finite limit as  $p_j \rightarrow \infty$  while if  $r_j > 1$  this tends to zero as  $p_j \rightarrow \infty$ .  $\square$

*Proof of Proposition 3.1.6.* Under the hypothesis, there exists  $p_j > \bar{p}_j$  such that  $p_j e^{u_j(p_j)} \geq (p_j)^{1-r} e^{\kappa + v_j}$ . Taking  $p_j \rightarrow \infty$ ,  $p_j e^{u_j(p_j)} \rightarrow \infty$  since  $r < 1$ . Clearly then  $p_j P_j^L(p_j, \mathbf{p}_{-j}) \rightarrow \infty$ . The claim follows.  $\square$

*Proof of Proposition 3.1.7.* The claim for  $\vartheta > -\infty$  is a consequence of Proposition 3.1.5. The claim for  $\vartheta = -\infty$  and  $\mathbf{p}_{-f} = \infty$  follows from Lemma 3.1.2, (v). If  $\vartheta = -\infty$  and  $\mathbf{p}_{-f} \neq \infty$ , then some competitor's product acts like an outside good for firm  $f$  and the proof is completed in the same way as the claim for  $\vartheta > -\infty$ .  $\square$

*Proof of Proposition 3.1.9.* This follows from Eqn. (3.6), observing that that  $(Dw_j)(p_j) P_j^L(\mathbf{p}) \neq 0$  for  $p_j < \infty$ .  $\square$

*Proof of Proposition 3.1.8.* We can write

$$\begin{aligned} (\nabla_f \hat{\pi}_f)(\mathbf{p}) &= (D_f \mathbf{w}_f)(\mathbf{p}_f) \text{diag}(\mathbf{P}_f^L(\mathbf{p})) \\ &\quad \times \left( (\mathbf{I} - \mathbf{1P}_f^L(\mathbf{p})^\top)(\mathbf{p}_f - \mathbf{c}_f) + (D_f \mathbf{w}_f)(\mathbf{p}_f)^{-1} \mathbf{1} \right). \end{aligned}$$

Stationarity then requires  $(\mathbf{I} - \mathbf{1P}_f^L(\mathbf{p})^\top)(\mathbf{p}_f - \mathbf{c}_f) + (D_f \mathbf{w}_f)(\mathbf{p}_f)^{-1} \mathbf{1} = \mathbf{0}$ . The Sherman-Morrison formula for the inverse of a rank-one perturbation of the identity (Ortega and Rheinboldt, 1970, Chapter 2, pg. 50) implies that

$$(\mathbf{I} - \mathbf{1P}_f^L(\mathbf{p})^\top)^{-1} = \mathbf{I} + \left( \frac{1}{1 - \mathbf{P}_f^L(\mathbf{p})^\top \mathbf{1}} \right) \mathbf{1P}_f^L(\mathbf{p})^\top; \quad (\text{D.1})$$

so long as  $\mathbf{P}_f^L(\mathbf{p})^\top \mathbf{1} < 1$ . This last condition will hold if either  $\vartheta > -\infty$  or, if  $\vartheta = -\infty$ ,  $\mathbf{p}_{-f} \neq \infty$ . Now if  $\vartheta = -\infty$  and  $\mathbf{p}_{-f} = \infty$ , no finite prices maximize  $\hat{\pi}_f$ , and hence we ignore this case. The claim follows.  $\square$

*Proof of Proposition 3.1.10.* To prove (i), we first prove that profit-maximizing prices are positive. No  $\mathbf{p}_f$  with non-positive profits can be a best response to  $\mathbf{p}_{-f}$ , since any  $\mathbf{p}_f > \mathbf{c}_f$  yields positive profits. So suppose  $\mathbf{p}_f$  is a best response to  $\mathbf{p}_{-f}$ , with  $p_j = 0$ .  $(D_j \hat{\pi}_j)(\mathbf{p} + \varepsilon \mathbf{e}_j)$  has the same sign as  $\hat{\pi}_f(\mathbf{p} + \varepsilon \mathbf{e}_j) - \sigma(Dw_j)(\varepsilon)^{-1} + c_j - \varepsilon$  for all  $\varepsilon > 0$ . Since  $\hat{\pi}$  is continuous at  $\mathbf{p}$  and  $\hat{\pi}(\mathbf{p}) > 0$ , there exists some  $\varepsilon^*$  so that  $\hat{\pi}(\mathbf{p} + \varepsilon \mathbf{e}_j) > 0$  if  $\varepsilon < \varepsilon^*$ . But then  $(D_j \hat{\pi}_j)(\mathbf{p} + \varepsilon \mathbf{e}_j)$  is positive for  $\varepsilon < \min\{c_j, \varepsilon^*\}$ , since  $\hat{\pi}_f(\mathbf{p} + \varepsilon \mathbf{e}_j) - \sigma(Dw_j)(\varepsilon)^{-1} + c_j > c_j$ . Thus profits *increase* with  $p_j$  near zero, contrary to the assumption that  $\mathbf{p}_f$  is a best response to  $\mathbf{p}_{-f}$ .

Now if  $\mathbf{p}_f \in [0, \infty)$  locally maximizes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$ , Eqn. (3.5) holds. However  $(\mathbf{I} - \mathbf{1P}_f^L(\mathbf{p})^\top)^{-1}$  maps positive vectors to positive vectors, a fact proved by the formula Eqn. (D.1). Since  $-(D\mathbf{w}_f)(\mathbf{p}_f)\mathbf{1} > \mathbf{0}$ , Eqn. (3.5) proves the claim. These two results prove (i), which allows  $p_j = \infty$  for some  $j \in \mathcal{J}_f$ .

To prove (ii), we observe that Proposition 3.1.7 rules out  $p_j = \infty$  for some  $j \in \mathcal{J}_f$ .  $\square$

*Proof of Lemma 3.1.13.* We observe the following bound, valid for large enough  $p_j$ :

$$p_j - c_j - \zeta_j(\mathbf{p}) \geq \left( \frac{r_j - 1}{r_j} \right) p_j - (c_j + \hat{\pi}_f(\mathbf{p}_f, \mathbf{p}_{-f})).$$

Under the hypothesis of (i),  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  is bounded (since  $w$  is eventually log bounded). Because  $r_j > 1$ ,  $((r_j - 1)/r_j)p_j \rightarrow \infty$  as  $p_j \rightarrow \infty$ . Thus we can always choose  $p_j$  large enough to make  $p_j - c_j - \zeta_j(\mathbf{p}) > 0$ . When  $\vartheta > -\infty$ ,  $\hat{\pi}_f(\cdot)$  itself is bounded and hence

$\bar{\mathbf{p}}_f$  can be chosen independently of  $\mathbf{p}_{-f}$ . The counter-claim is just as straightforward: Suppose  $(Dw_j)(p_j) \geq -p_j^{-1}$  for all  $p_j > \bar{p}_j$ . Then  $(Dw_j)(p_j)^{-1} \leq -p_j$  and hence

$$p_j - c_j - \zeta_j(\mathbf{p}) \leq p_j - p_j - (c_j + \hat{\pi}_f(\mathbf{p})) \leq 0$$

if  $p_j > \bar{p}_j$ . □

*Proof of Theorem 3.1.12.* Claim (i) follows from the following componentwise formula for the intra-firm profit price-Hessians  $(D_f \nabla_f \hat{\pi}_f)(\mathbf{p})$ :

$$\begin{aligned} (D_l D_k \hat{\pi}_{f(k)})(\mathbf{p}) &= \delta_{k,l} (Dw_k)(p_k) P_k^L(\mathbf{p}) \\ &\quad \times \left( \left( \frac{(D^2 w_k)(p_k) + (Dw_k)(p_k)^2}{(Dw_k)(p_k)} \right) (p_k - c_k - \hat{\pi}_{f(k)}(\mathbf{p})) + 2 \right) \\ &\quad - (D_k \hat{\pi}_{f(k)})(\mathbf{p}) P_l^L(\mathbf{p}) (Dw_l)(p_l) \\ &\quad - (Dw_k)(p_k) P_k^L(\mathbf{p}) (D_l \hat{\pi}_{f(l)})(\mathbf{p}) \end{aligned} \tag{D.2}$$

This formula is a consequence of Proposition (C.2.1). Particularly, when  $\mathbf{p}_f$  makes  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  stationary,

$$(D_f \nabla_f \hat{\pi}_f)(\mathbf{p}) = (D_f \mathbf{w}_f)(\mathbf{p}_f) \text{diag}(\mathbf{P}_f(\mathbf{p})) (\mathbf{I} - \mathbf{\Omega}_f(\mathbf{p}_f))$$

where  $\mathbf{\Omega}_f(\mathbf{p}_f)$  is a diagonal matrix with entries

$$\omega_j = \frac{(D^2 w_j)(p_j)}{(Dw_j)(p_j)^2}.$$

Thus, the Hessians are diagonal matrices with negative diagonal entries when  $w$  has sub-quadratic second derivatives.

Claim (ii) is an obvious corollary to (i). Claim (iii) follows from the Poincare-Hopf Theorem applied to the (continuous) negative gradient vector field  $-(\nabla_f \hat{\pi}_f)(\cdot, \mathbf{p}_{-f})$  on  $[\mathbf{c}_f, \bar{\mathbf{p}}_f(\mathbf{p}_{-f})]$ , with the first part of Lemma 3.1.13 instrumental in proving that the negative gradient vector field points outward on the boundary of  $[\mathbf{c}_f, \bar{\mathbf{p}}_f(\mathbf{p}_{-f})]$ . The Poincare-Hopf Theorem states that the sum of the indices of the negative gradient vector field at all stationary points equals one, the Euler characteristic of  $[\mathbf{c}_f, \bar{\mathbf{p}}_f(\mathbf{p}_{-f})]$ . But by (i), all stationary points have index equal to one because they are minimizers of the function generating the gradient vector field (Milnor, 1965). Hence there can only be a single stationary point. Claim (iv) is a corollary to the third, and does not require the assumption that  $\vartheta > -\infty$ . However, it does require stipulating that there

is a finite solution to the fixed-point equation.  $\square$

*Proof of Theorem 3.1.14.* Let  $\bar{\mathbf{p}}$  be as in Lemma 3.1.13, (ii).  $\Xi(\mathbf{p}) = \mathbf{p} - \mathbf{c} - \zeta(\mathbf{p})$  is a vector field on  $[\mathbf{c}, \bar{\mathbf{p}}]$  that points outward on  $[\mathbf{c}, \bar{\mathbf{p}}] \setminus [\mathbf{c}, \bar{\mathbf{p}}]$ . Since  $\zeta_j(\mathbf{p}) > 0$  when  $[\mathbf{c}, \bar{\mathbf{p}}] \setminus (\mathbf{c}, \bar{\mathbf{p}}]$ ,  $\Xi(\mathbf{p})$  points outward on  $[\mathbf{c}, \bar{\mathbf{p}}] \setminus (\mathbf{c}, \bar{\mathbf{p}}]$ , and hence on all of the boundary of  $[\mathbf{c}, \bar{\mathbf{p}}]$ . Let the set of zeros of  $\Xi$  be denoted by  $\mathfrak{Z} = \{\mathbf{p} \in (\mathbf{c}, \bar{\mathbf{p}}) : \Xi(\mathbf{p}) = \mathbf{0}\}$  and let  $\iota_{\Xi}(\mathbf{p})$  denote the index of  $\Xi$  at  $\mathbf{p} \in \mathfrak{Z}$ . The Poincare-Hopf Theorem states that  $\sum_{\mathbf{p} \in \mathfrak{Z}} \iota_{\Xi}(\mathbf{p}) = 1$ , where the value of the sum on the left is taken to be 0 if  $\mathfrak{Z} = \{\emptyset\}$ . Hence  $|\mathfrak{Z}| > 0$ .  $\square$

*Proof of Proposition 3.1.16.* Eqn. (D.2) states that

$$\begin{aligned} & (D_l D_k \hat{\pi}_f)(\mathbf{q}) \\ &= |(Dw_k)(q_k)| P_k^L(\mathbf{q}) (\hat{\pi}_f(\mathbf{q}) - (q_k - c_k) - (Dw_k)(q_k)^{-1}) P_l^L(\mathbf{q}) |(Dw_l)(q_l)| \\ & \quad + |(Dw_k)(q_k)| P_k^L(\mathbf{q}) (\hat{\pi}_f(\mathbf{q}) - (q_l - c_l) - (Dw_l)(q_l)^{-1}) P_l^L(\mathbf{q}) |(Dw_l)(q_l)| \end{aligned}$$

for any  $\mathbf{q}$ . Our goal is to choose  $\mathbf{q}$ ,  $\|\mathbf{q} - \mathbf{p}\| < \varepsilon$ , so that  $\hat{\pi}_f(\mathbf{q}) - (q_k - c_k) - (Dw_k)(q_k)^{-1} < 0$  and  $\hat{\pi}_f(\mathbf{q}) - (q_l - c_l) - (Dw_l)(q_l)^{-1}$ .

Let  $k, l \in \mathcal{J}_f$ ,  $k \neq l$ , and choose  $\delta > 0$  so that  $\mathbf{p}_f$  defined by  $p_j = p_j^*$  if  $j \in \mathcal{J}_f \setminus \{k, l\}$  and  $p_j = p_j^* + \delta$  otherwise also satisfies  $\|\mathbf{p} - \mathbf{p}^*\| < \varepsilon$ . Assume that  $\hat{\pi}_f(\cdot, \mathbf{p}_{-f})$  is maximized by  $\mathbf{p}_f^*$ . By our fixed-point characterization,

$$\begin{aligned} & \hat{\pi}_f(\mathbf{p}_f, \mathbf{p}_{-f}) - (p_k - c_k) - (Dw_k)(p_k)^{-1} \\ & < \hat{\pi}_f(\mathbf{p}_f^*, \mathbf{p}_{-f}) - (p_k - c_k) - (Dw_k)(p_k)^{-1} + \delta \\ & = \hat{\pi}_f(\mathbf{p}_f^*, \mathbf{p}_{-f}) - (p_k^* - c_k) - (Dw_k)(p_k)^{-1} \\ & = (Dw_k)(p_k^*)^{-1} - (Dw_k)(p_k)^{-1}. \end{aligned}$$

When  $w$  is concave,

$$(Dw_k)(p_k^*)^{-1} - (Dw_k)(p_k)^{-1} = (Dw_k)(p_k^*)^{-1} - (Dw_k)(p_k^* + \delta)^{-1} \leq 0$$

proving that  $\hat{\pi}_f(\mathbf{p}_f, \mathbf{p}_{-f}) - (Dw_k)(p_k^*)^{-1} - (p_k - c_k) < 0$ . The same argument applies to  $l$  and thus  $(D_l D_k \hat{\pi}_f)(\mathbf{p}) < 0$ , where  $\mathbf{p} = (\mathbf{p}_f, \mathbf{p}_{-f})$ .

For the second claim, note that

$$(D_l D_k \log \hat{\pi}_f)(\mathbf{p}) = \frac{(D_l D_k \hat{\pi}_f)(\mathbf{p}) \hat{\pi}_f(\mathbf{p}) - (D_k \hat{\pi}_f)(\mathbf{p}) (D_l \hat{\pi}_f)(\mathbf{p})}{\hat{\pi}_f(\mathbf{p})^2}.$$

We have already established that the first term in the numerator is negative at  $\mathbf{p}$  as defined above. Furthermore,

$$(D_k \hat{\pi}_f)(\mathbf{p}) = |(Dw_k)(p_k)| P_k^L(\mathbf{p})(\hat{\pi}_f(\mathbf{p}) - (p_k - c_k) - (Dw_k)(p_k)^{-1}) < 0$$

by the same argument and hence  $(D_k \hat{\pi}_f)(\mathbf{p})(D_l \hat{\pi}_f)(\mathbf{p}) > 0$ , making the second term in the numerator also negative. This completes the proof.  $\square$

### Intra-Firm Structure

*Proof of Corollary 3.2.1.* This follows immediately from the fixed-point expression  $\mathbf{p}_f = \mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p})$  which is equivalent to the markup equation  $p_j - c_j = \hat{\pi}_f(\mathbf{p}) - \sigma(Dw_j)(p_j)^{-1}$  for all  $j \in \mathcal{J}_f$ .  $\square$

*Proof of Lemma 3.2.3:* Quite simply,  $(D^p \varphi)(\mathbf{y}, p) = 1 - \omega(\mathbf{y}, p)$  where  $\omega(\mathbf{y}, p) = (D^2 w)(\mathbf{y}, p)/(Dw)(\mathbf{y}, p)^2$ .  $\square$

*Proof of Corollary 3.2.4.* Let  $\varphi(p) = p - (Dw_j)(p)^{-1} = p - (Dw_k)(p)^{-1}$  for any  $p \in [0, \infty)$ . Because  $c_j = c_k$ , Eqn. (3.9) implies that  $\varphi(p_j) = \varphi(p_k)$ . But  $\varphi$  is strictly increasing, and hence one-to-one, by Lemma 3.2.3, and thus  $p_j = p_k$ .  $\square$

*Proof of Corollary 3.2.5.* We prove that  $p_j - c_j \geq p_k - c_k$  implies  $c_j \leq c_k$ . By Corollary 3.2.1,  $p_j - c_j \geq p_k - c_k$  implies  $(Dw)(p_j)^{-1} \leq (Dw)(p_k)^{-1}$ , or equivalently  $(Dw)(p_j) \geq (Dw)(p_k)$ . By strict concavity, this implies that  $p_j \leq p_k$ . But then  $c_j - c_k \leq p_j - p_k \leq 0$ .  $\square$

*Proof of Corollary 3.2.6.* Note that  $c_j \leq c_k$  implies  $v_j \leq v_k$  is equivalent to the value costs hypothesis. Then the claim follows from Corollary 3.2.5.  $\square$

*Proof of Corollary 3.2.7.* The unique value hypothesis implies that when  $v(\mathbf{y}) > v(\mathbf{y}')$ ,  $\varphi(\mathbf{y}, p) \leq \varphi(\mathbf{y}', p)$  for all  $p \in [0, \infty)$ . Specifically, if  $v(\mathbf{y}_j) > v(\mathbf{y}_k)$  then  $\varphi(\mathbf{y}_j, p_k) \leq \varphi(\mathbf{y}_k, p_k)$ . Suppose that  $v(\mathbf{y}_j) > v(\mathbf{y}_k)$  while  $p_j \leq p_k$ . Because  $\varphi(\mathbf{y}_j, p)$  is a strictly increasing function of  $p$ , we have

$$\varphi(\mathbf{y}_j, p_j) \leq \varphi(\mathbf{y}_j, p_k) \leq \varphi(\mathbf{y}_k, p_k).$$

Thus Eqn. (3.9) implies that  $c_j - c_k = \varphi(\mathbf{y}_j, p_j) - \varphi(\mathbf{y}_k, p_k) \leq 0$  in contradiction to the value costs hypothesis.  $\square$

## Inter-Firm Structure

*Proof of Corollary 3.2.10.* Simply observe that  $\psi(\mathbf{y}_j, \cdot) = \psi(\mathbf{y}_k, \cdot)$  is strictly increasing, so that  $\psi(\mathbf{y}_j, p) - \psi(\mathbf{y}_k, q) > 0$  if, and only if,  $p > q$  and so on.  $\square$

## D.2 Proofs for Chapter 4

*Proof of Proposition 4.1.1:* This follows directly from the Leibniz Rule and the following form of the derivatives of the Logit choice probabilities:

$$(D_k P_j^L)(\boldsymbol{\theta}, \mathbf{p}) = P_j^L(\boldsymbol{\theta}, \mathbf{p})(\delta_{j,k} - P_k^L(\boldsymbol{\theta}, \mathbf{p}))(Dw_k)(\boldsymbol{\theta}, p_k).$$

$\square$

*Proof of Proposition 4.1.2:* We first note that

$$|(D_k P_j^L)(\boldsymbol{\theta}, \mathbf{p})| \leq |(Dw_k)(\boldsymbol{\theta}, p_k)| e^{u_k(\boldsymbol{\theta}, p_k) - \vartheta(\boldsymbol{\theta})}$$

regardless of  $j, k$ . The continuous differentiability of  $P_j$  assuming the Leibniz Rule and the uniform  $\mu$ -integrability hypothesis is a direct result of this bound and the Dominated Convergence Theorem. That the Leibniz Rule follows from this bound and the uniform  $\mu$ -integrability hypothesis is a consequence of the Mean Value Theorem and the Dominated Convergence Theorem. Particularly, the mean value theorem for functions of a single real variable states that

$$h^{-1}(P_j^L(\boldsymbol{\theta}, \mathbf{p} + h\mathbf{e}_k) - P_j^L(\boldsymbol{\theta}, \mathbf{p})) = (D_k P_j^L)(\boldsymbol{\theta}, \mathbf{p} + \eta\mathbf{e}_k)$$

for some  $\eta$  such that  $|\eta| < |h|$ . Thus

$$h^{-1} |P_j^L(\boldsymbol{\theta}, \mathbf{p} + h\mathbf{e}_k) - P_j^L(\boldsymbol{\theta}, \mathbf{p})| \leq |(Dw_k)(\boldsymbol{\theta}, p_k + \eta)| e^{u_k(\boldsymbol{\theta}, p_k + \eta) - \vartheta(\boldsymbol{\theta})} \leq \varphi_{(\mathbf{y}_k, p_k)}(\boldsymbol{\theta})$$

for  $\mu$ -a.e.  $\boldsymbol{\theta} \in \mathcal{T}$  and the Dominated Convergence Theorem validates the Leibniz Rule. This proof is essentially that given in a general setting by (Bartle, 1966, Chapter 5, pg. 46).  $\square$

**A Remark.** An “easier” bound is simply  $|(D_k P_j^L)(\boldsymbol{\theta}, \mathbf{p})| \leq |(Dw_k)(\boldsymbol{\theta}, p_k)|$ , and thus we might consider changing the statement of Proposition 4.1.2 to hypothesize only the uniform  $\mu$ -integrability of the utility price derivatives. In fact, we use this below to validate the fixed-point approach for the Boyd and Mellman (1980) model



that lacks an outside good. However, this bound fails to be useful for a central utility specification of the form (e.g. [Berry et al. \(1995\)](#); [Petrin \(2002\)](#))

$$w(\boldsymbol{\theta}, \mathbf{y}, p) = \alpha(\boldsymbol{\theta}) \log(\varsigma(\boldsymbol{\theta}) - p)$$

for which  $|(Dw_k)(\boldsymbol{\theta}, p_k)| = \alpha(\mathbf{y})/(\varsigma(\boldsymbol{\theta}) - p)$  is singular on  $\varsigma^{-1}(p)$ . In empirical applications,  $\varsigma$  is onto, generating a singularity somewhere in  $\mathcal{T}$  for all  $p$ . Thus this singularity cannot be “controlled” for all  $p$  by choosing the measure  $\mu$ . Thus a hypothesis on only the utility price derivatives is not useful. We show below that our hypothesis including the exponentiated utility function can be used to prove the validity of the fixed-point equation under this same utility specification.

*Proof of Proposition 4.1.4:* (i) follows easily from the defining relationships.

(ii) If  $(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \mathbf{0}$ , then  $\boldsymbol{\zeta}(\mathbf{p}) = \boldsymbol{\eta}(\mathbf{p}) = \mathbf{p} - \mathbf{c}$ . For the opposite implication, we use the formulae just derived. The first formula implies that if  $\boldsymbol{\zeta}(\mathbf{p}) = \boldsymbol{\eta}(\mathbf{p})$ , then  $\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p}) \in \mathcal{N}(\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top)$ . The second formula is equivalent to

$$\begin{aligned} \mathbf{p} - \mathbf{c} - \boldsymbol{\eta}(\mathbf{p}) &= (\mathbf{I} - \boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top)^{-1}(\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p})) \\ &= \sum_{n=1}^{\infty} (\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top)^n (\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p})), \end{aligned}$$

where the equality

$$(\mathbf{I} - \boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top)^{-1} = \sum_{n=1}^{\infty} (\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top)^n$$

follows from the observation that  $\rho(\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top) \leq \|\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top\|_\infty < 1$  ([Lemma 4.1.5](#), Claim (i)). Thus  $\mathbf{p} - \mathbf{c} - \boldsymbol{\zeta}(\mathbf{p}) \in \mathcal{N}(\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top) \subset \mathcal{N}((\boldsymbol{\Lambda}(\mathbf{p})^{-1}\tilde{\boldsymbol{\Gamma}}(\mathbf{p})^\top)^n)$  implies that  $\mathbf{p} = \mathbf{c} + \boldsymbol{\eta}(\mathbf{p})$ , in turn implying that  $(\tilde{\nabla}\hat{\pi})(\mathbf{p}) = \mathbf{0}$ .  $\square$

*Proof of Lemma 4.1.5:* (i): We note that

$$(\boldsymbol{\Lambda}_f(\mathbf{p})^{-1}\boldsymbol{\Gamma}_f(\mathbf{p})^\top)_{k,l} = \frac{\gamma_{l,k}(\mathbf{p})}{\lambda_k(\mathbf{p})} = \int P_l^L(\boldsymbol{\theta}, \mathbf{p}) d\mu_{k,\mathbf{p}}(\boldsymbol{\theta})$$

where  $\mu_{k,\mathbf{p}}$  is the probability distribution with density, with respect to  $\mu$ , given by

$$d\mu_{k,\mathbf{p}}(\boldsymbol{\theta}) = \frac{P_k^L(\boldsymbol{\theta}, \mathbf{p}) |(Dw_k)(\boldsymbol{\theta}, p_k)| d\mu(\boldsymbol{\theta})}{\int P_k^L(\boldsymbol{\phi}, \mathbf{p}) |(Dw_k)(\boldsymbol{\phi}, p_k)| d\mu(\boldsymbol{\phi})}.$$

Thus  $\mathbf{\Lambda}_f(\mathbf{p})^{-1}\mathbf{\Gamma}_f(\mathbf{p})^\top$  has row sums

$$\int \left( \sum_{j \in \mathcal{J}_f} P_j^L(\boldsymbol{\theta}, \mathbf{p}) \right) d\mu_{k, \mathbf{p}}(\boldsymbol{\theta}) < 1.$$

The additional assumption that  $\vartheta : \mathcal{T} \rightarrow (-\infty, \infty)$  plays a role in establishing this inequality because then there is always a set  $\mathcal{T}'_k \subset \mathcal{T}$  with  $\mu_{k, \mathbf{p}}(\mathcal{T}'_k) > 0$  on which  $\sum_{j \in \mathcal{J}_f} P_j^L(\boldsymbol{\theta}, \mathbf{p}) < 1$ . Our claim follows.

(ii): The inequality

$$1 > \int \left( \sum_{j \in \mathcal{J}_f} P_j^L(\boldsymbol{\theta}, \mathbf{p}) \right) d\mu_{k, \mathbf{p}}(\boldsymbol{\theta})$$

is equivalent to

$$\begin{aligned} \left| (\mathbf{I} - \mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top)_{k, k} \right| &= 1 - \int P_k^L(\boldsymbol{\theta}, \mathbf{p}) d\mu_{k, \mathbf{p}}(\boldsymbol{\theta}) \\ &> \int \left( \sum_{j \in \mathcal{J}_f \setminus k} P_j^L(\boldsymbol{\theta}, \mathbf{p}) \right) d\mu_{k, \mathbf{p}}(\boldsymbol{\theta}) \\ &= \sum_{l \neq k} \left| (\mathbf{I} - \mathbf{\Lambda}(\mathbf{p})^{-1}\tilde{\mathbf{\Gamma}}(\mathbf{p})^\top)_{k, l} \right|. \end{aligned}$$

The claim follows.

(iii): Because

$$(\mathbf{I} - \mathbf{\Lambda}_f(\mathbf{p})^{-1}\mathbf{\Gamma}_f(\mathbf{p})^\top)^{-1} = \sum_{n=1}^{\infty} (\mathbf{\Lambda}_f(\mathbf{p})^{-1}\mathbf{\Gamma}_f(\mathbf{p})^\top)^n$$

and  $\mathbf{\Lambda}_f(\mathbf{p})^{-1}\mathbf{\Gamma}_f(\mathbf{p})^\top$  maps positive vectors to positive vectors, so does its power series; i.e., so does  $(\mathbf{I} - \mathbf{\Lambda}_f(\mathbf{p})^{-1}\mathbf{\Gamma}_f(\mathbf{p})^\top)^{-1}$ .

(iv): Let  $x_k = \max \mathbf{x}$ ,  $x_l = \min \mathbf{x} < 0$ , and  $\{\omega_j\}_{j \in \mathcal{J}_f} \subset [0, 1]$  be any set of numbers that sum to less than one. Then

$$\sum_{j \in \mathcal{J}_f} \omega_j x_j \leq \left( \sum_{j \in \mathcal{J}_f} \omega_j \right) x_k < x_k \quad \text{and} \quad \sum_{j \in \mathcal{J}_f} \omega_j x_j \geq \left( \sum_{j \in \mathcal{J}_f} \omega_j \right) x_l > x_l,$$

the very last inequality requiring the assumption that  $x_l < 0$ . Taking, respectively,

$$\omega_j = (\Lambda_f(\mathbf{p})^{-1}\Gamma_f(\mathbf{p})^\top)_{k,j} \quad \text{and} \quad \omega_j = (\Lambda_f(\mathbf{p})^{-1}\Gamma_f(\mathbf{p})^\top)_{l,j},$$

we obtain

$$((\mathbf{I} - \Lambda_f(\mathbf{p})^{-1}\Gamma_f(\mathbf{p})^\top)\mathbf{x})_k > 0 \quad \text{and} \quad ((\mathbf{I} - \Lambda_f(\mathbf{p})^{-1}\Gamma_f(\mathbf{p})^\top)\mathbf{x})_l < 0.$$

□

*Proof of Corollary 4.1.6:* This follows from Eqn. (4.1) and Lemma 4.1.5, (iii), since  $-\Lambda_f(\mathbf{p})\mathbf{P}_f(\mathbf{p}) > \mathbf{0}$ . □

*Proof of Proposition 4.3.1:* The [Boyd and Mellman \(1980\)](#) model takes

$$\begin{aligned} w(\boldsymbol{\theta}, \mathbf{y}, p) &= -\theta_{K+1}p \\ v(\boldsymbol{\theta}, \mathbf{y}) &= \sum_{k=1}^K \theta_k y_k \end{aligned}$$

with no outside good. As mentioned above,  $|(D_k P_j^L)(\boldsymbol{\theta}, \mathbf{p})| \leq |(Dw_k)(\boldsymbol{\theta}, p_k)|$  always holds. While this is not useful in general, it can be applied here to derive  $|(D_k P_j^L)(\boldsymbol{\theta}, \mathbf{p})| \leq \theta_{K+1}$ , a bound that demonstrates the uniform  $\mu$ -integrability of the choice probability derivatives for any probability measure  $\mu$  giving a finite expected price coefficient. Thus the conclusions of Proposition 4.1.2 hold, since there is a finite expected price coefficient under the [Boyd and Mellman](#) model.

The [Berry et al. \(1995\)](#) model takes  $\boldsymbol{\theta} = (\phi, \boldsymbol{\psi})$  and

$$\begin{aligned} w(\phi, \mathbf{y}, p) &= \alpha \log(\phi - p) \\ v(\boldsymbol{\psi}, \mathbf{y}) &= \sum_{k=1}^K \psi_k \gamma_k y_k \\ \vartheta(\phi, \boldsymbol{\psi}) &= \alpha \log(\phi) + \gamma_0 \psi_0 \end{aligned}$$

The hypothesis of Proposition 4.1.2 is not satisfied if  $\alpha < 1$ : For if  $\phi > p$ , we have

$$|(Dw)(\boldsymbol{\theta}, \mathbf{y}, p)| e^{u(\boldsymbol{\theta}, \mathbf{y}, p) - \vartheta(\boldsymbol{\theta})} = \alpha \left(\frac{1}{\phi}\right) \left(\frac{1}{\phi - p}\right)^{1-\alpha} e^{v(\boldsymbol{\psi}, \mathbf{y})}$$

which is not  $\mu$ -integrable at  $\phi = p$ , for any  $p > 0$ , for any probability measure  $\mu$

supported on a subset of  $[0, \infty)$ . By assuming  $\alpha > 1$  we have instead

$$|(Dw)(\boldsymbol{\theta}, \mathbf{y}, p)| e^{u(\boldsymbol{\theta}, \mathbf{y}, p) - \vartheta(\boldsymbol{\theta})} = \alpha \left( \frac{1}{\phi} \right) (\phi - p)^{\alpha-1} e^{v(\boldsymbol{\psi}, \mathbf{y})}$$

which  $\mu$ -integrable for any probability measure  $\mu$  supported on a subset of  $[0, \infty)$  that vanishes suitably quickly at zero. In addition,  $|(Dw)(\boldsymbol{\theta}, \mathbf{y}, p)| e^{u(\boldsymbol{\theta}, \mathbf{y}, p) - \vartheta(\boldsymbol{\theta})}$  vanishes as  $\phi \downarrow p$ . Thus we set

$$|(Dw)(\boldsymbol{\theta}, \mathbf{y}, p)| e^{u(\boldsymbol{\theta}, \mathbf{y}, p) - \vartheta(\boldsymbol{\theta})} = \varphi_p(\phi) e^{-\gamma_0 \psi_0 + \sum_{k=1}^K \psi_k \gamma_k y_k}$$

where now

$$\varphi_p(\phi) = \begin{cases} \alpha \left( \frac{(\phi - p)^{\alpha-1}}{\phi} \right) & \text{if } \phi > p \\ 0 & \text{if } \phi \leq p \end{cases}$$

We can take  $\mu = \mu_p \times \mu_y$  where  $\mu_p$  is a measure for  $\phi$  giving it a lognormal distribution and  $\mu_y$  is a measure for  $\boldsymbol{\psi}$  giving it a normal distribution. By Fubini's theorem,

$$\begin{aligned} & \int |(Dw)(\boldsymbol{\theta}, \mathbf{y}, p)| e^{u(\boldsymbol{\theta}, \mathbf{y}, p) - \vartheta(\boldsymbol{\theta})} d\mu(\boldsymbol{\theta}) \\ &= \left( \int e^{-\gamma_0 \psi_0 + \sum_{k=1}^K \psi_k \gamma_k y_k} d\mu_y(\boldsymbol{\psi}) \right) \int \varphi_p(\phi) d\mu_p(\phi) \end{aligned}$$

The normal distribution on  $\boldsymbol{\psi}$  ensures that

$$\int e^{-\gamma_0 \psi_0 + \sum_{k=1}^K \psi_k \gamma_k y_k} d\mu_y(\boldsymbol{\psi})$$

is finite. We turn to the price term,  $\int \varphi_p(\phi) d\mu_p(\phi)$ . If  $p \leq p'$ , then  $\varphi_p(\phi) \geq \varphi_{p'}(\phi)$  for all  $\phi$ . Thus,  $\varphi_p(\phi) \leq \varphi_0(\phi)$  for all  $\phi$  for any  $p > 0$ . Now  $\varphi_0(\phi) = \alpha \phi^{\alpha-2}$ , an integrable function with respect to the measure  $\mu_p$  for any  $\alpha > 1$ . Hence  $|(Dw)(\boldsymbol{\theta}, \mathbf{y}, p)| e^{u(\boldsymbol{\theta}, \mathbf{y}, p) - \vartheta(\boldsymbol{\theta})}$  is uniformly  $\mu$ -integrable in a neighborhood of  $p$ , for any  $p > 0$ .  $\square$

## D.3 Proofs for Chapter 5

### Regulatory Policies

*Proof of Proposition 5.2.1:* These formula are derived by setting  $\sum_{f=1}^F \hat{c}_f^R(\mathbf{p}) = 0$  and solving for  $\bar{\kappa}^S(\mathbf{p})$ . For example,

$$0 = \sum_{f=1}^F \hat{c}_f^R(\mathbf{p}) = \mathbf{P}(\mathbf{p})^\top \boldsymbol{\kappa} - \kappa^S (\mathbf{P}(\mathbf{p})^\top \mathbf{1}).$$

□

### Smooth Stationarity

*Proof of Proposition 5.3.1:* This is a straightforward application of the chain rule. □

*Proof of Proposition 5.3.2:* Given Eqn. (5.2) and the fact that no diagonal element of  $\boldsymbol{\Lambda}_f(\mathbf{p})$  is zero, this formula just algebra. □

### Incentives

*Proof of Proposition 5.3.3:* It is easy to see that

$$\begin{aligned} (\nabla_f \hat{\pi}_f^R)(\mathbf{p}) &= (\nabla_f \hat{\pi}_f)(\mathbf{p}) \\ &\quad + r_f [|\boldsymbol{\Lambda}_f(\mathbf{p})| (\mathbf{I} - \boldsymbol{\Lambda}_f(\mathbf{p})^{-1} \boldsymbol{\Gamma}_f(\mathbf{p})^\top) \boldsymbol{\varrho}_f(\mathbf{p}) - (D_f \boldsymbol{\varrho}_f)(\mathbf{p})^\top \mathbf{P}_f(\mathbf{p})] \end{aligned}$$

where  $|\mathbf{A}|$  is the matrix the same size as  $\mathbf{A}$  with entries given by the absolute value of the entries of  $\mathbf{A}$ , not the determinant of  $\mathbf{A}$ . For CO<sub>2</sub> taxes and fixed-pivot feebates,  $(D_f \boldsymbol{\varrho}_f)(\mathbf{p}) \equiv \mathbf{0}$  and hence

$$(\nabla_f \hat{\pi}_f^R)(\mathbf{p}) = (\nabla_f \hat{\pi}_f)(\mathbf{p}) + r_f |\boldsymbol{\Lambda}_f(\mathbf{p})| (\mathbf{I} - \boldsymbol{\Lambda}_f(\mathbf{p})^{-1} \boldsymbol{\Gamma}_f(\mathbf{p})^\top) \boldsymbol{\varrho}_f(\mathbf{p}).$$

Thus if  $(\nabla_f \hat{\pi}_f)(\mathbf{p}) = \mathbf{0}$ ,

$$(\nabla_f \hat{\pi}_f^R)(\mathbf{p}) = r_f |\boldsymbol{\Lambda}_f(\mathbf{p})| (\mathbf{I} - \boldsymbol{\Lambda}_f(\mathbf{p})^{-1} \boldsymbol{\Gamma}_f(\mathbf{p})^\top) \boldsymbol{\varrho}_f(\mathbf{p}).$$

Recalling Lemma 4.1.5, Claim (iv), we obtain (i), (ii), and (iii).

For (iv) and (v), we note that  $(D_f \boldsymbol{\varrho}_f)(\mathbf{p})^\top \mathbf{P}_f(\mathbf{p}) = (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})(\nabla_f \hat{\varrho}_f)(\mathbf{p})$  because  $\boldsymbol{\varrho}_f = \hat{\varrho}_f \mathbf{1}$  for some function  $\hat{\varrho}_f : \mathbb{R}^J \rightarrow \mathbb{R}$ . Writing  $(\nabla_f \hat{\varrho}_f)(\mathbf{p}) = (D_f \mathbf{P}_f)(\mathbf{p})^\top \boldsymbol{\chi}_f(\mathbf{p})$  as

in the text we obtain

$$(\nabla_f \hat{\pi}_f^R)(\mathbf{p}) = r_f |\boldsymbol{\Lambda}_f(\mathbf{p})| (\mathbf{I} - \boldsymbol{\Lambda}_f(\mathbf{p})^{-1} \boldsymbol{\Gamma}_f(\mathbf{p})^\top) [\hat{\varrho}_f(\mathbf{p}) \mathbf{1} + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_f(\mathbf{p})]$$

when  $(\nabla_f \hat{\pi}_f)(\mathbf{p}) = \mathbf{0}$ . Consider the CAE standard. The formula given in Proposition 5.3.6, Claim (iii) proves that

$$\hat{\varrho}_f(\mathbf{p}) \mathbf{1} + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_f(\mathbf{p}) = \kappa_f^S \mathbf{1} - \boldsymbol{\kappa}_f.$$

Claim (v) follows. Now consider the CAFE standard. The formula given in Proposition 5.3.6, Claim (i) proves that

$$\hat{\varrho}_f(\mathbf{p}) \mathbf{1} + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_f(\mathbf{p}) = \bar{\kappa}_f^H(\mathbf{p}) \left[ \left( \frac{\kappa_f^S}{\bar{\kappa}_f^H(\mathbf{p})} \right) \mathbf{1} - \frac{\bar{\kappa}_f^H(\mathbf{p})}{\boldsymbol{\kappa}_f} \right]$$

Note that

$$\begin{aligned} & (\hat{\varrho}_f(\mathbf{p}) \mathbf{1} + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_f(\mathbf{p}))_k - (\hat{\varrho}_f(\mathbf{p}) \mathbf{1} + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_f(\mathbf{p}))_l \\ &= \bar{\kappa}_f^H(\mathbf{p})^2 \left( \frac{1}{\kappa_l} - \frac{1}{\kappa_k} \right) > 0 \end{aligned}$$

if, and only if,  $\kappa_k < \kappa_l$ . Thus the *maximal* element of  $\hat{\varrho}_f(\mathbf{p}) \mathbf{1} + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_f(\mathbf{p})$  is determined by the *minimal* element of  $\boldsymbol{\kappa}_f$ , and vice versa.

Next, we recognize that

$$\hat{\varrho}_f(\mathbf{p}) + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_j(\mathbf{p}) = \kappa_f^S - 2\bar{\kappa}_f^H(\mathbf{p}) + \frac{\bar{\kappa}_f^H(\mathbf{p})^2}{\kappa_j} > 0$$

if and only if

$$\frac{\kappa_f^S}{\bar{\kappa}_f^H(\mathbf{p})} - 1 > 1 - \frac{\bar{\kappa}_f^H(\mathbf{p})}{\kappa_j}.$$

Because  $\kappa_f^S > \bar{\kappa}_f^H(\mathbf{p})$ , the left hand side is positive, while for any  $\kappa_j < \bar{\kappa}_f^H(\mathbf{p})$  the right hand side is negative. Thus  $\hat{\varrho}_f(\mathbf{p}) + (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\chi}_j(\mathbf{p}) > 0$  for the vehicle  $j$  with minimal  $\kappa_j$ . Thus the maximal element of  $\hat{\varrho}_f(\mathbf{p}) \mathbf{1} - \boldsymbol{\chi}_f(\mathbf{p})$  is positive. Claim (iv) follows. □

## Non-Smooth Stationarity

*Proof of Proposition 5.3.4:* We prove the result for a monopolist firm (that faces an outside good). For fixed competitor prices, these problems are equivalent, though as competitors change prices the value of the outside good changes. Thus, in what follows we neglect the “ $f$ ” subscript.

This problem falls into the framework studied by Ioffe (1979), who considered minimization problems over some Banach space  $\mathbb{X}$  with objective  $\phi(\mathbf{x}) = g(\varphi(\mathbf{x}))$  where  $\varphi : \mathbb{X} \rightarrow \mathbb{Y}$ ,  $\mathbb{Y}$  another Banach space, is continuously differentiable and  $g : \mathbb{Y} \rightarrow \mathbb{R}$  is sublinear. Take  $\mathbb{X} = \mathbb{R}^J$ ,  $\mathbb{Y} = \mathbb{R}^2$ ,

$$\varphi(\mathbf{p}) = \left( \hat{\pi}(\mathbf{p}), (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\hat{\rho}(\mathbf{p}) \right),$$

and  $g(x, y) = -x + r \max\{0, y\}$ ;  $g$  is sublinear because its epigraph is a nonempty convex cone Hiriart-Urruty and Lemarechal (2001). These choices casts the maximization of  $\hat{\pi}^R$  in the class of those characterized by Ioffe.

Ioffe (1979) defines the Lagrangian

$$\ell_{(\lambda, \mu)}(\mathbf{p}) = (\lambda, \mu)^\top \varphi(\mathbf{p}) = \lambda \hat{\pi}(\mathbf{p}) + \mu (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\hat{\rho}(\mathbf{p})$$

The first order necessary conditions are that  $\mathbf{0} \in (\partial\phi)(\mathbf{p})$  (Clarke, 1975), where  $(\partial\phi)(\mathbf{p})$  is the generalized gradient of  $\phi$ ; i.e. that there is a  $\boldsymbol{\lambda} \in (\partial g)(\varphi(\mathbf{p}))$  such that  $(D\varphi)(\mathbf{p})^\top \boldsymbol{\lambda} = \mathbf{0}$  (Ioffe, 1979). Since

$$(\partial g)(x, y) = \begin{cases} \{(-1, 0)\} & y < 0 \\ \{-1\} \times [0, r] & y = 0 \\ \{(-1, r)\} & y > 0 \end{cases}$$

and

$$(D\varphi)(\mathbf{p})^\top = \begin{bmatrix} | & | \\ (\nabla \hat{\pi})(\mathbf{p}) & \nabla[(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\hat{\rho}(\mathbf{p})] \\ | & | \end{bmatrix}$$

this amounts to the existence of  $\lambda_* \in [0, r]$  such that  $(\nabla \hat{\pi})(\mathbf{p}) = \lambda_* \nabla[(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\hat{\rho}(\mathbf{p})]$ .

Let

$$\Omega = \{\boldsymbol{\lambda} \in (\partial g)(\varphi(\mathbf{p})) \subset \mathbb{R}^2 : (D\varphi)(\mathbf{p})^\top \boldsymbol{\lambda} = \mathbf{0}\}$$

Ioffe (1979) gives the following condition:

**Theorem D.3.1.** (Ioffe, 1979) *If  $\phi$  attains a local minimum at  $\mathbf{p}$ , then, in addition*

to the first order conditions,

$$\max\{\mathbf{q}^\top (D\nabla\ell_\lambda)(\mathbf{p})\mathbf{q} : \lambda \in \Omega\} \geq 0$$

for all  $\mathbf{q} \in \mathbb{R}^J$  satisfying  $g(\boldsymbol{\varphi}(\mathbf{p})) + (D\boldsymbol{\varphi})(\mathbf{p})\mathbf{q} \leq g(\boldsymbol{\varphi}(\mathbf{p}))$ .

Moreover, if in addition to the first-order conditions there is a  $\mu > 0$  such that

$$\max_{\lambda \in \Omega} \{\mathbf{q}^\top (D\nabla\ell_\lambda)(\mathbf{p})\mathbf{q}\} \geq \mu \|\mathbf{q}\|_2^2 \quad (\text{D.3})$$

for all

$$\mathbf{q} \in \mathcal{K} = \{\mathbf{h} : g(\boldsymbol{\varphi}(\mathbf{p})) + t(D\boldsymbol{\varphi})(\mathbf{p})\mathbf{h} = g(\boldsymbol{\varphi}(\mathbf{p})) \text{ for some } t > 0\}$$

then  $\mathbf{p}$  is a local minimum of  $\phi$ .

For us,  $(\nabla\ell_\lambda)(\mathbf{p}) = \lambda_1(\nabla\hat{\pi})(\mathbf{p}) + \lambda_2(\nabla\hat{\rho})(\mathbf{p})$  and thus

$$(D\nabla\ell_\lambda)(\mathbf{p}) = \lambda_1(D\nabla\hat{\pi})(\mathbf{p}) + \lambda_2(D\nabla)[(\mathbf{P}(\mathbf{p})^\top \mathbf{1})\hat{\rho}(\mathbf{p})].$$

Furthermore,  $\Omega$  is a singleton,  $\boldsymbol{\lambda}_* = (-1, \lambda_*)$ , where  $\lambda_* < 1$  is given by the first-order necessary conditions. Thus,

$$\max_{\lambda \in \Omega} \{\mathbf{q}^\top (D\nabla\ell_\lambda)(\mathbf{p})\mathbf{q}\} = \mathbf{q}^\top (D\nabla\ell_{(-1, \lambda_*)})(\mathbf{p})\mathbf{q}.$$

Now we characterize  $\mathcal{K}$  when  $\hat{\rho}(\mathbf{p}) = 0$ .

**Lemma D.3.2.** (i) If  $\lambda_* \in (0, r)$ , then  $\mathcal{K} = (\nabla\hat{\rho})(\mathbf{p})^\perp$ . (ii) If  $\lambda_* = 0$ , then  $\mathcal{K} = \{\mathbf{h} : (\nabla\hat{\rho})(\mathbf{p})^\top \mathbf{h} \leq 0\}$ . (iii) If  $\lambda_* = r$ , then  $\mathcal{K} = \{\mathbf{h} : (\nabla\hat{\rho})(\mathbf{p})^\top \mathbf{h} \geq 0\}$ .

*Proof.* First note that  $\nabla[(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\hat{\rho}(\mathbf{p})] = (\nabla\hat{\rho})(\mathbf{p})$  when  $\hat{\rho}(\mathbf{p}) = 0$ .

Next,  $g(\boldsymbol{\varphi}(\mathbf{p})) = g(\hat{\pi}(\mathbf{p}), 0) = -\hat{\pi}(\mathbf{p})$  and

$$(D\boldsymbol{\varphi})(\mathbf{p}) = \begin{bmatrix} \lambda_* \\ 1 \end{bmatrix} (\nabla\hat{\rho})(\mathbf{p})^\top$$

Thus

$$(D\boldsymbol{\varphi})(\mathbf{p})\mathbf{h} = ((\nabla\hat{\rho})(\mathbf{p})^\top \mathbf{h}) \begin{bmatrix} \lambda_* \\ 1 \end{bmatrix}$$



and

$$\boldsymbol{\varphi}(\mathbf{p}) + t(D\boldsymbol{\varphi})(\mathbf{p})\mathbf{h} = \begin{bmatrix} \hat{\pi}(\mathbf{p}) + t\lambda_*((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) \\ t((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) \end{bmatrix},$$

which implies

$$g(\boldsymbol{\varphi}(\mathbf{p}) + t(D\boldsymbol{\varphi})(\mathbf{p})\mathbf{h}) - g(\boldsymbol{\varphi}(\mathbf{p})) = -t\lambda_*((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) + tr \begin{cases} 0 & \text{if } ((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) \leq 0 \\ ((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) & \text{if } ((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) > 0 \end{cases}$$

If  $(\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h} \leq 0$  then  $\mathbf{h} \in \mathcal{K}$  if and only if  $tr\lambda_*((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) = 0$ . If  $(\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h} \geq 0$  then  $\mathbf{h} \in \mathcal{K}$  if and only if  $t(r - \lambda_*)((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) = 0$ .

To prove (i), we note that since  $\lambda_* \in (0, r)$ ,  $\mathbf{h} \in \mathcal{K}$  if and only if  $\mathbf{h} \perp \nabla[(\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})\hat{\rho}(\mathbf{p})]$ . To prove (ii) and (iii), note that the appropriate inequalities are trivially satisfied.  $\square$

To apply the second order necessary condition, we also consider when  $g(\boldsymbol{\varphi}(\mathbf{p}) + (D\boldsymbol{\varphi})(\mathbf{p})\mathbf{h}) \leq g(\boldsymbol{\varphi}(\mathbf{p}))$ .

**Lemma D.3.3.**  $\{\mathbf{h} : g(\boldsymbol{\varphi}(\mathbf{p}) + (D\boldsymbol{\varphi})(\mathbf{p})\mathbf{h}) \leq g(\boldsymbol{\varphi}(\mathbf{p}))\} = \mathcal{K}$ .

*Proof.* Clearly  $g(\boldsymbol{\varphi}(\mathbf{p}) + (D\boldsymbol{\varphi})(\mathbf{p})\mathbf{h}) \leq g(\boldsymbol{\varphi}(\mathbf{p}))$  holds for all  $\mathbf{h} \perp \nabla[(\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})\hat{\rho}(\mathbf{p})] = (\nabla\hat{\rho})(\mathbf{p})$ . If  $(\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h} < 0$ , we must have  $-r\lambda_*(\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h} \leq 0$  which cannot be satisfied unless  $\lambda_* = 0$  or  $\mathbf{h} \perp (\nabla\hat{\rho})(\mathbf{p})$ . If  $(\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h} > 0$  then we must have  $(r - \lambda_*)((\nabla\hat{\rho})(\mathbf{p})^\top\mathbf{h}) \leq 0$ , which again cannot be satisfied unless  $\lambda_* = r$  or  $\mathbf{h} \perp (\nabla\hat{\rho})(\mathbf{p})$ .  $\square$

To complete the proof, we note that

$$(D\nabla\ell_{(-1, \lambda_*)})(\mathbf{p}) = -(D\nabla\hat{\pi})(\mathbf{p}) + \lambda_*(D\nabla)[(\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})\hat{\rho}(\mathbf{p})]$$

and compute  $\mathbf{q}^\top(D\nabla)[(\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})\hat{\rho}(\mathbf{p})]\mathbf{q}$  on  $(\nabla\hat{\rho})(\mathbf{p})^\perp$ .

**Lemma D.3.4.** *We have that*

$$\nabla[(\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})\hat{\rho}(\mathbf{p})] = \hat{\rho}(\mathbf{p})((D_f\mathbf{P}_f)(\mathbf{p})^\top\mathbf{1}) + (\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})(\nabla\hat{\rho})(\mathbf{p})$$

and, when  $\hat{\rho}(\mathbf{p}) = 0$ ,

$$(D\nabla)[(\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})\hat{\rho}(\mathbf{p})] = ((D_f\mathbf{P}_f)(\mathbf{p})^\top\mathbf{1})(\nabla\hat{\rho})(\mathbf{p})^\top + (\nabla\hat{\rho})(\mathbf{p})((D_f\mathbf{P}_f)(\mathbf{p})^\top\mathbf{1})^\top + (\mathbf{P}_f(\mathbf{p})^\top\mathbf{1})(D\nabla\hat{\rho})(\mathbf{p}).$$

In particular,

$$\mathbf{q}^\top (D\nabla)[(\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\hat{\rho}(\mathbf{p})]\mathbf{q} = (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})\mathbf{q}^\top (D\nabla\hat{\rho})(\mathbf{p})\mathbf{q}$$

when  $\hat{\rho}(\mathbf{p}) = 0$  and  $\mathbf{q} \perp (\nabla\hat{\rho})(\mathbf{p})$ .

*Proof.* These are relatively simple calculations.  $\square$

Ioffe (1979)'s second order necessary condition stipulates that if  $\mathbf{p}$  is a local minimizer of  $\phi$ , then  $(D\nabla\ell_{(-1,\lambda_*)})(\mathbf{p})$  is positive semi-definite on  $\mathcal{K}$ . By the previous Lemma, this is equivalent to  $(D\nabla\hat{\pi})(\mathbf{p}) - \lambda_*(D\nabla\hat{\rho})(\mathbf{p})$  negative semi-definite on  $\mathcal{K}$ . Similarly the second order sufficient condition holds if and only if  $(D\nabla\ell_{(-1,\lambda_*)})(\mathbf{p})$  is positive definite on  $\mathcal{K}$ ; that is, if  $(D\nabla\hat{\pi})(\mathbf{p}) - \lambda_*(D\nabla\hat{\rho})(\mathbf{p})$  is negative definite on  $(\nabla\hat{\rho})(\mathbf{p})^\perp$ . These are the claims in Proposition 5.3.4.  $\square$

## Shadow Taxes

*Proof of Proposition 5.3.5:* The first-order conditions are

$$(\nabla_f\hat{\pi}_f)(\mathbf{p}) = v_f(\nabla_f\hat{\rho}_f)(\mathbf{p}) = v_f(D_f\mathbf{P}_f)(\mathbf{p})^\top \boldsymbol{\chi}_f(\mathbf{p}),$$

equivalent to

$$\begin{aligned} \mathbf{p}_f &= \mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) + v_f(\mathbf{I} - \boldsymbol{\Lambda}_f(\mathbf{p})^{-1}\boldsymbol{\Gamma}_f(\mathbf{p})^\top) \boldsymbol{\chi}_f(\mathbf{p}) \\ &= \mathbf{c}_f + \boldsymbol{\Lambda}_f(\mathbf{p})^{-1}(\boldsymbol{\Gamma}_f(\mathbf{p})^\top(\mathbf{p}_f - \mathbf{c}_f) - \mathbf{P}_f(\mathbf{p})) \\ &\quad + v_f\boldsymbol{\chi}_f(\mathbf{p}) - v_f\boldsymbol{\Lambda}_f(\mathbf{p})^{-1}\boldsymbol{\Gamma}_f(\mathbf{p})^\top \boldsymbol{\chi}_f(\mathbf{p}) \\ &= \mathbf{c}_f + v_f\boldsymbol{\chi}_f(\mathbf{p}) + \boldsymbol{\Lambda}_f(\mathbf{p})^{-1}(\boldsymbol{\Gamma}_f(\mathbf{p})^\top(\mathbf{p}_f - \mathbf{c}_f - v_f\boldsymbol{\chi}_f(\mathbf{p})) - \mathbf{P}_f(\mathbf{p})) \\ &= \mathbf{c}_f + \mathbf{t}_f + \boldsymbol{\Lambda}_f(\mathbf{p})^{-1}(\boldsymbol{\Gamma}_f(\mathbf{p})^\top(\mathbf{p}_f - \mathbf{c}_f - \mathbf{t}_f) - \mathbf{P}_f(\mathbf{p})) \end{aligned}$$

Clearly then the first order conditions for the flat tax/subsidy problem are satisfied at  $\mathbf{p}_f$ .  $\square$

## Hybrid Fixed-Point Iteration

*Proof of Proposition 5.4.1:* We first prove that  $(\nabla_f\hat{\pi}_f)(\mathbf{p})$  and

$$\lim_{\mathbf{q}_f \rightarrow \mathbf{p}_f} (\nabla_f\hat{\pi}_f^R)(\mathbf{q}_f, \mathbf{p}_{-f}),$$

where the limit is taken over any sequence  $\{\mathbf{q}_f^{(n)}\}_{n \in \mathbb{N}}$  satisfying  $\hat{\varrho}_f(\mathbf{q}_f^{(n)}, \mathbf{p}_{-f}) > 0$ , are colinear and point in opposite directions. Note that  $(\nabla_f \hat{\pi}_f)(\mathbf{p}) = v_f (\nabla_f \hat{\varrho}_f)(\mathbf{p})$  for some  $v_f < r_f (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1})$ . Now we write

$$\begin{aligned} (\nabla_f \hat{\pi}_f^R)(\mathbf{q}_f, \mathbf{p}_{-f}) &= (\nabla_f \hat{\pi}_f)(\mathbf{q}_f, \mathbf{p}_{-f}) \\ &\quad - r_f \left( \hat{\varrho}_f(\mathbf{q}_f, \mathbf{p}_{-f}) (D_f \mathbf{P}_f)(\mathbf{q}_f, \mathbf{p}_{-f})^\top \mathbf{1} \right. \\ &\quad \left. + (\mathbf{P}_f(\mathbf{q}_f, \mathbf{p}_{-f})^\top \mathbf{1}) (\nabla_f \hat{\varrho}_f)(\mathbf{q}_f, \mathbf{p}_{-f}) \right) \end{aligned}$$

for  $\hat{\varrho}_f(\mathbf{p}_f, \mathbf{p}_{-f}) > 0$ . Note that as  $\mathbf{q}_f \rightarrow \mathbf{p}_f$  (maintaining  $\hat{\varrho}_f(\mathbf{p}_f, \mathbf{p}_{-f}) > 0$ )

$$\begin{aligned} (\nabla_f \hat{\pi}_f^R)(\mathbf{q}_f, \mathbf{p}_{-f}) &\rightarrow (\nabla_f \hat{\pi}_f)(\mathbf{p}) - r_f (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) (\nabla_f \hat{\varrho}_f)(\mathbf{p}) \\ &= -(r_f (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) - v_f) (\nabla_f \hat{\varrho}_f)(\mathbf{p}) \\ &= - \left( \frac{r_f (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) - v_f}{v_f} \right) (\nabla_f \hat{\pi}_f)(\mathbf{p}) \end{aligned}$$

In other words, with  $\alpha_f^R = (r_f (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) - v_f) / v_f > 0$  we have

$$\lim_{\mathbf{q}_f \rightarrow \mathbf{p}_f} (\nabla_f \hat{\pi}_f^R)(\mathbf{q}_f, \mathbf{p}_{-f}) = -\alpha_f^R (\nabla_f \hat{\pi}_f)(\mathbf{p}).$$

We now complete the proof. First recall that  $\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) - \mathbf{p}_f = -\boldsymbol{\Lambda}_f(\mathbf{p}) (\nabla_f \hat{\pi}_f)(\mathbf{p})$  and hence

$$\begin{aligned} \mathbf{c}_f + \boldsymbol{\zeta}_f^R(\mathbf{p}) + r_f (\mathbf{P}_f(\mathbf{p})^\top \mathbf{1}) \boldsymbol{\Lambda}_f(\mathbf{p})^{-1} (\nabla_f \hat{\varrho}_f)(\mathbf{p}) - \mathbf{p}_f \\ &= -\boldsymbol{\Lambda}_f(\mathbf{p}) \lim_{\mathbf{q}_f \rightarrow \mathbf{p}_f} (\nabla_f \hat{\pi}_f^R)(\mathbf{q}_f, \mathbf{p}_{-f}) \\ &= -\alpha_f^R (-\boldsymbol{\Lambda}_f(\mathbf{p}) (\nabla_f \hat{\pi}_f)(\mathbf{p})) \\ &= -\alpha_f^R (\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) - \mathbf{p}_f). \end{aligned}$$

Finally, note that

$$\begin{aligned} (\mathbf{c}_f + \boldsymbol{\zeta}_f(\mathbf{p}) - \mathbf{p}_f)^\top (\nabla_f \hat{\varrho}_f)(\mathbf{p}) &= (\nabla_f \hat{\pi}_f)(\mathbf{p})^\top (-\boldsymbol{\Lambda}_f(\mathbf{p})^{-1}) (\nabla_f \hat{\varrho}_f)(\mathbf{p}) \\ &= v_f (\nabla_f \hat{\varrho}_f)(\mathbf{p})^\top (-\boldsymbol{\Lambda}_f(\mathbf{p})^{-1}) (\nabla_f \hat{\varrho}_f)(\mathbf{p}) > 0 \end{aligned}$$

(because  $\boldsymbol{\Lambda}_f(\mathbf{p})$  has only negative entries). □

# Bibliography

- Allenby, Greg M, Peter E Rossi. 1991. Quality perceptions and asymmetric switching between brands. *Marketing Science* **10**(3) 185–204.
- Anderson, Simon P., Andre de Palma. 1992a. The logit as a model of product differentiation. *Oxford Economic Papers* **44**(1) 51–67.
- Anderson, Simon P, Andre de Palma. 1992b. Multiproduct firms: A nested logit approach. *The Journal of Industrial Economics* **40**(3) 271–276.
- Anderson, Simon P., Andre de Palma. 2001. Product diversity in asymmetric oligopoly: Is the quality of consumer goods too low? *The Journal of Industrial Economics* **49**(2) 113–135.
- Anderson, Simon P., Andre de Palma. 2006. Market performance of multi-product firms. *The Journal of Industrial Economics* **54**(1) 95–124.
- Anderson, Simon P., Andre de Palma, Yurii Nesterov. 1995. Oligopolistic competition and the optimal provision of products. *Econometrica* **63**(6) 1281–1301.
- Austin, David, Terry Dinan. 2005. Clearing the air: The costs and consequences of higher caf standards and increased gasoline taxes. *Journal of Environmental Economics and Management* **50** 562–582.
- AVL. 2008a. Advisor software. URL <http://www.avl.com>.
- AVL. 2008b. Cruise software. URL <http://www.avl.com>.
- Bajari, Patrick, C. Lanier Benkard, Jonathan Levin. 2007. Estimating dynamic models of imperfect competition. *Econometrica* **75**(5) 1331–1370.
- Bartle, Robert G. 1966. *The Elements of Integration and Lebesgue Measure*. Jon Wiley and Sons, Inc.
- Baye, Michael R., Dan Kovenock. 2008. Bertrand competition. Steven N. Durlaf, Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics, Second Edition*. Palgrave Macmillian.
- Benkard, C. Lanier. 2004. A dynamic analysis of the market for wide-bodied commercial aircraft. *Review of Economic Studies* **71** 581–611.
- Bento, Antonio M., Lawrence M. Goulder, Emeric Henry, Mark R. Jacobsen, Roger H. von Haefen. 2005. Distributional and efficiency impacts of gasoline taxes: An econometrically based multi-market study. *The American Economic Review* **95**(2) 282–287.
- Beresteanu, Arie, Shanjun Li. 2008. Gasoline prices, government support, and the demand for hybrid vehicles in the u.s. Working Paper.

- Berndt, Ernst R., Ann F. Friedlander, Judy Sahw-Er Wang Chiang. 1990. Interdependent pricing and markup behavior: An empirical analysis of gm, ford, and chrysler. Tech. Rep. 3396, National Bureau of Economic Research (NBER).
- Berry, Steven, James Levinsohn, Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–890.
- Berry, Steven, James Levinsohn, Ariel Pakes. 2004. Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy* **112**(1) 68–105.
- Besanko, David, Sachin Gupta, Dipak Jain. 1998. Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science* **44**(11) 1533–1547.
- Bierlaire, Michel, Denis Bolduc, Daniel McFadden. 2003. Characteristics of generalized extreme value distributions. Tech. rep., University of California, Berkeley.
- BLAS. 2008. Basic linear algebra subroutines. URL <http://www.netlib.org/blas/>.
- Boyd, J Hayden, Robert E Mellman. 1980. The effect of fuel economy standards on the u.s. automotive market: An hedonic demand analysis. *Transportation Research A* **14A** 367–378.
- Bresnahan, Timothy. 1987. Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics* **35**(4) 457–482.
- Brown, Peter N., Youcef Saad. 1990. Hybrid krylov methods for nonlinear systems of equations. *SIAM Journal of Scientific and Statistical Computing* **11**(3) 450–481.
- Broyden, C. G., J. E. Dennis Jr., Jorge J. More. 1973. On the local and superlinear convergence of quasi-newton methods. *Journal of the Institute for Mathematics and its Applications* **12** 223–245.
- Byrd, Richard H., Jorge Nocedal, Richard A. Waltz. 2006. Knitro: An integrated package for nonlinear optimization. Tech. rep., Ziena Inc. URL [www.ziena.com/papers/integratedpackage](http://www.ziena.com/papers/integratedpackage).
- Caplin, Andrew, Barry Nalebuff. 1991. Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica* **59**(1) 25–59.
- CARB. 2008. Climate change for mobile sources. URL <http://www.arb.ca.gov/cc/ccms/ccms.htm>.
- CBO. 2003. The economic costs of fuel economy standards versus a gasoline tax. Tech. rep., Congressional Budget Office.
- Choi, S Chan, Wayne S Desarbo, Patrick T Harker. 1990. Product positioning under price competition. *Management Science* **36**(2) 175–199.

- Clarke, Frank H. 1975. Generalized gradients and applications. *Transactions of the American Mathematical Society* **205** 247–262.
- CPS. 2007. Current population survey. URL <http://www.census.gov/cps/>.
- d’Aspremont, C, J Jaskold Gabszewicz, J-F Thisse. 1979. On hotelling’s ”stability in competition”. *Econometrica* **47**(5) 1145–1150.
- Dembo, Ron S., Stanley C. Eisenstat, Trond Steihaug. 1982. Inexact newton methods. *SIAM Journal on Numerical Analysis* **19**(2) 400–408.
- Dennis, John E., Robert B. Schnabel. 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM.
- Dingell, John. 2007. Dingellcast 14 - climate change. URL <http://www.house.gov/dingell/Podcasts/Dingellcast14.htm>.
- Doraszelski, Ulrich, Michaela Draganska. 2006. Market segmentation strategies of multiproduct firms. *Journal of Industrial Economics* **54**(1) 125–149.
- Doyle, Maura P. 1997. The effects of interest rates and taxes on new car prices. Tech. Rep. 1997-38, Federal Reserve Bank.
- Draganska, Michaela, Dipak Jain. 2004. A likelihood approach to estimating market equilibrium models. *Management Science* **50**(5) 605–616.
- Economides, Nicholas S. 1987. On nash equilibrium existence and optimality in oligopolistic competition in prices and varieties. *Greek Economic Review* **9**(2) 198–209.
- EIA. 2003. The national energy modeling system: An overview. URL <http://www.eia.doe.gov/oiaf/aeo/overview/index.html>.
- EIA. 2007. Transportation sector module of the national energy modeling system: Model documentation 2007. Tech. Rep. DOE/EIA-M070(2007), Department of Energy.
- EIA. 2008a. Assumptions to the annual energy outlook 2008. URL <http://www.eia.doe.gov/oiaf/aeo/assumption/index.html>.
- EIA. 2008b. Responses to congressional and other requests. URL [http://www.eia.doe.gov/oiaf/service\\_rpts.htm](http://www.eia.doe.gov/oiaf/service_rpts.htm).
- Eisenstat, Stanley C., Homer F. Walker. 1994. Globally convergent inexact newton methods. *SIAM Journal on Optimization* **4**(2) 393–422.
- Eisenstat, Stanley C., Homer F. Walker. 1996. Choosing the forcing terms in an inexact newton method. *SIAM Journal of Scientific Computing* **17**(1) 16–32.

- Ericson, Richard, Ariel Pakes. 1995. Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies* **62** 53–82.
- Feller, William. 1968. *Introduction to Probability and its Applications, Volume I*. Wiley.
- Fershtman, Chaim, Neil Gandal, Sarit Markovich. 1999. Estimating the effect of tax reform in differentiated product oligopolistic markets. *Journal of Public Economics* **74** 151–170.
- Fudenberg, Drew, Jean Tirole. 1991. *Game Theory*. MIT Press.
- GM. 2008. Introducing the 2009 escalade hybrid. URL <http://www.cadillac.com/escaladehybrid/>.
- Goldberg, Pinelopi K. 1995. Product differentiation and oligopoly in international markets: The case of the u.s. automobile industry. *Econometrica* **63**(4) 891–951.
- Goldberg, Pinelopi K. 1998. The effects of the corporate average fuel efficiency standards in the us. *The Journal of Industrial Economics* **46**(1) 1–33.
- Golub, Gene H, Charles F Van Loan. 1996. *Matrix Computations*. The Johns Hopkins University Press.
- Gowrisankaran, Gautam, Robert J. Town. 1997. Dynamic equilibrium in the hospital industry. *Journal of Economics and Management Strategy* **6**(1) 45–74.
- Greene, David L., John DeCicco. 2000. Engineering-economic analyses of automotive fuel economy potential in the united states. *Annual Review of Energy and the Environment* **25** 477–536.
- Greene, David L., Philip D. Patterson, Margaret Singh, Jia Li. 2005. Feebates, rebates and gas-guzzler taxes: A study of incentives for increased fuel economy. *Energy Policy* **33**(6) 757–775.
- Grimmett, Geoffrey R., David R. Stirzaker. 2001. *Probability and Random Processes*. Oxford University Press.
- Hanson, Ward, Kipp Martin. 1996. Optimizing multinomial logit profit functions. *Management Science* **42**(7) 992–1003.
- Hill, Kim, Morgan Edwards, Steven Szalsky. 2007. How automakers plan their products. Tech. rep., Center for Automotive Research.
- Hiriart-Urruty, Jean-Baptiste, Claude Lemarechal. 2001. *Fundamentals of Convex Analysis*. Springer-Verlag.
- Hotelling, Harold. 1929. Stability in competition. *The Economic Journal* **39**(153) 41–57.



- IATA. 2008. Jet fuel price monitor. URL <http://www.iata.org/>.
- Ioffe, A. D. 1979. Necessary and sufficient conditions for a local minimum, 3: Second-order conditions and augmented duality. *SIAM Journal on Control and Optimization* **17** 266–288.
- Jacobsen, Mark R. 2006. Evaluating u. s. fuel economy standards in a model with producer and household heterogeneity. Job Market Paper (Stanford University).
- Johnson, L. W., D. R. Scholz. 1968. On steffensen’s method. *SIAM Journal of Numerical Analysis* **5**(2) 296–302.
- Kakutani, Shizuo. 1941. A generalization of brouwer’s fixed point theorem. *Duke Mathematics Journal* **8**(3) 457–459.
- Kellogg, R. B. 1975. Uniqueness in the schauder fixed point theorem. *Proceedings of the American Mathematical Society* **60**(1) 207–210.
- KINSOL. 2008. Description of kinsol. URL [https://computation.llnl.gov/casc/sundials/description/description.html#descr\\_kinsol](https://computation.llnl.gov/casc/sundials/description/description.html#descr_kinsol).
- Kleit, Andrew. 1990. The effect of annual changes in automobile fuel economy standards. *Journal of Regulatory Economics* **2**(2) 151–172.
- Kleit, Andrew. 2004. Impacts of long-range increases in the corporate average fuel economy (cafe) standard. *Economic Inquiry* **42**(2) 279–294.
- KNITRO. 2008. Knitro. URL <http://www.ziena.com/knitro.htm>.
- Levin, Carl. 2007a. Levin statement on cafe deal. URL <http://levin.senate.gov/newsroom/release.cfm?id=288119>.
- Levin, Carl. 2007b. Statement of senator carl levin at the committee on commerce, science and transportation hearing on corporate average fuel economy (cafe) legislation. URL <http://levin.senate.gov/newsroom/release.cfm?id=273547>.
- Louvierre, Jordan J., David A. Hensher, Joffre D. Swait. 2000. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press.
- Luce, R. Duncan, Howard Raiffa. 1957. *Games and Decisions: A Critical Survey*. Jon Wiley and Sons, Inc.
- Luo, Zhi-Quan, Jong-Shi Pang, Daniel Ralph. 1996. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press.
- Maynard, Micheline. 2007. E.p.a. denies california emissions waiver.
- McFadden, Daniel, Kenneth Train. 2000. Mixed mnl models for discrete response. *Journal of Applied Econometrics* **15**(5) 447–440.

- McFadden, Daniel L. 1981. Econometric models of probabilistic choice. Charles F. Manski, Daniel L. McFadden, eds., *Structural Analysis of Discrete Data and Econometric Applications*. MIT Press, 199–272.
- McFadden, Daniel L. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57**(5) 995–1026.
- Michalek, Jeremy J, Panos Y Papalambros, Steven J Skerlos. 2004. A study of fuel efficiency and emission policy impact on optimal vehicle design decisions. *Journal of Mechanical Design* **126**(6) 1062–1070.
- Milgrom, Paul, John Roberts. 1990. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica* **58**(6) 1255–1277.
- Milnor, John W. 1965. *Topology from the Differentiable Viewpoint*. Princeton Landmarks in Mathematics, Princeton University Press.
- MINPACK. 2008. URL <http://www.netlib.org/minpack/>.
- Mizuno, Toshihide. 2003. On the existence of a unique price equilibrium for models of product differentiation. *International Journal of Industrial Organization* **21** 761–793.
- More, J. J., J. A. Trangenstein. 1976. On the global convergence of broyden’s method. *Mathematics of Computation* **30**(1) 523–540.
- Munkres, James R. 1991. *Analysis on Manifolds*. Westview Press.
- Myerson, Roger B. 1997. *Game Theory: Analysis of Conflict*. Harvard University Press.
- Nagle, Thomas, Reed Holden. 1987. *The Strategy and Tactics of Pricing*. Prentice Hall.
- Nash, John F. 1950. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences* **36** 48–49.
- Nash, John F. 1951. Non-cooperative games. *Annals of Mathematics* **54**(2) 286–295.
- NESCCAF. 2004. Reducing greenhouse gas emissions from light-duty vehicles. Tech. rep., Northeast States Center for a Clean Air Future.
- Nevo, Aviv. 2000a. Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics* **31**(3) 395–421.
- Nevo, Aviv. 2000b. A practitioner’s guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy* **9**(4) 513–548.
- Nevo, Aviv. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2) 307–342.

- NHTSA. 2006. Light truck average fuel economy standards, model years 2008-2011. Tech. rep., National Highway Traffic Safety Administration.
- NHTSA. 2008a. Average fuel economy standards, passenger cars and light trucks, model years 2011-2015. Tech. rep., National Highway Traffic Safety Administration.
- NHTSA. 2008b. Corporate average fuel economy. URL <http://www.nhtsa.dot.gov/portal/site/nhtsa/menuitem.d0b5a45b55bfbe582f57529cdba046a0/>.
- NHTSA. 2008c. Revised summary of fuel economy performance. Tech. rep., National Highway Traffic Safety Administration.
- NHTSA. 2008d. Title 49 united states code, chapter 329-automotive fuel economy. URL <http://www.nhtsa.dot.gov/CARS/rules/CAFE/legislation.htm>.
- NIST/SEMATECH. 2008. Extreme value distributions. URL <http://www.itl.nist.gov/div898/handbook/apr/section1/apr163.htm>.
- Nocedal, Jorge, Dong C. Liu. 1989. On the limited memory method for large scale optimization. *Mathematical Programming B* **45**(3) 503–528.
- NRC. 2002. *The Effectiveness and Impact of the Corporate Average Fuel Economy Standards*. National Academies Press.
- Ortega, J. M., W. C. Rheinboldt. 1970. *Iterative Solution of Nonlinear Equations in Several Variables*. Society for Industrial and Applied Mathematics.
- Pakes, Ariel, Paul McGuire. 1994. Computing markov-perfect nash equilibria: Numerical implications of a dynamic differentiated product model. *RAND Journal of Economics* **25**(4) 555–589.
- Pavley, Fran. 2002. Assembly bill 1493. URL <http://www.arb.ca.gov/bluebook/bb07/POP/POP-AB1493.htm>.
- Pawłowski, Roger P., John N. Shadid, Joseph P. Simonis, Homer F. Walker. 2006. Globalization techniques for newton-krylov methods and applications to the fully coupled solutions of the navier-stokes equations. *SIAM Review* **48**(4) 700–721.
- Pawłowski, Roger P., Joseph P. Simonis, Homer F. Walker, John N. Shadid. 2008. Inexact newton dogleg methods. *SIAM Journal of Numerical Analysis* **46**(4) 2112–2132.
- Pernice, Michael, Homer F. Walker. 1998. Nitsol: A newton iterative solver for nonlinear systems. *SIAM Journal of Scientific Computing* **19**(1) 302–318.
- Petrin, Amil. 2002. Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy* **110**(4) 705–729.

- Powell, M. J. D. 1970. A hybrid method for nonlinear equations. Philip Rabinowitz, ed., *Numerical Methods for Nonlinear Algebraic Equations*, chap. 6. Gordon and Breach Science Publishers, 87–114.
- Rossi, Peter, Greg M Allenby, Robert McCulloch. 2006. *Bayesian Statistics and Marketing*. Wiley.
- Saad, Youcef, Martin Schultz. 1986. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal of Scientific and Statistical Computing* **7**(3) 856–869.
- Sandor, Zsolt. 2001. Computation, efficiency and endogeneity in discrete choice models. Ph.D. thesis, University of Groningen.
- Shaked, Avner, John Sutton. 1982. Relaxing price competition through product differentiation. *The Review of Economic Studies* **49**(1) 3–13.
- Shaked, Avner, John Sutton. 1990. Multiproduct firms and market structure. *The RAND Journal of Economics* **21**(1) 45–62.
- Shapiro, Carl. 1989. Theories of oligopoly behavior. R. Schmalensee, R. D. Willig, eds., *Handbook of Industrial Organization*, vol. I. Elsevier.
- Simsek, Alp, Asuman Ozdaglar, Daron Acemoglu. 2007. Generalized poincare-hopf theorem for compact nonsmooth regions. *Mathematics of Operations Research* **32**(1) 193–214.
- Skerlos, Steven J, Jeremy J Michalek, W Ross Morrow. 2005. Sustainable design engineering and science: Selected challenges and case studies. M. A. Abraham, ed., *Sustainable Science and Engineering, Volume 1: Defining Principles*, chap. 23. Elsevier.
- Stavins, Robert N. 2007. Addressing climate change with a comprehensive u.s. cap-and-trade system. *The Oxford Review of Economic Policy (Forthcoming)* URL [http://ksghome.harvard.edu/~rstavins/Papers/Stavins'\\_Article\\_on\\_US\\_Cap-and-Trade\\_for\\_Oxford\\_Review.pdf](http://ksghome.harvard.edu/~rstavins/Papers/Stavins'_Article_on_US_Cap-and-Trade_for_Oxford_Review.pdf).
- Sudhir, K. 2001. Competitive pricing behavior in the auto market: A structural analysis. *Marketing Science* **20**(1) 42–60.
- Train, Kenneth. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Train, Kenneth, Clifford Winston. 2007. Vehicle choice behavior and the declining market share of u.s. automakers. *International Economic Review* **48**(4) 1469–1496.
- Trefethen, Lloyd N., David Bau. 1997. *Numerical Linear Algebra*. SIAM.
- USAToday. 2008. Rising costs reshaping air travel across the usa.

- Verboven, Frank. 1996. International price discrimination in the european car market. *The RAND Journal of Economics* **27**(2) 240–268.
- Verboven, Frank. 1999. Product line rivalry and market segmentation - with an application to automobile optional engine pricing. *Journal of Industrial Economics* **47**(4) 399–425.
- Verboven, Frank. 2002. Quality-based price discrimination and tax incidence: Evidence from gasoline and diesel cars. *The RAND Journal of Economics* **33**(2) 275–297.
- Villas-Boas, J Miguel, Ying Zhao. 2005. Retailers, manufacturers, and individual consumers: Modeling the supply side in the ketchup marketplace. *Journal of Marketing Research* **41**(1) 83–95.
- Viswanath, Divakar. 2007. Recurrent motions within plane couette turbulence. *Journal of Fluid Dynamics* **580** 339–358. URL <http://arxiv.org/abs/physics/0604062>.
- Viswanath, Divakar. 2008. The critical layer in pipe flow at high reynolds number. (Personally Communicated).
- Walker, Homer F. 1988. Implementation of the gmres method using householder transformations. *SIAM Journal of Scientific and Statistical Computing* **9**(1) 152–163.
- Walker, Homer F. 2005. Nitsol read me. URL <http://users.wpi.edu/~walker/NITSOL/readme.txt>.
- Wards. 2004-2007. *Wards Automotive Yearbooks*. Wards Reports Inc.
- Weniger, Ernst Joachim. 1989. Nonlinear sequence transformations: Computational tools for the acceleration of convergence and the summation of divergent series. *Computer Physics Reports* **10**(5-6) 189–371. URL <http://arxiv.org/abs/math/0306302>.