

NBER WORKING PAPER SERIES

EMPIRICAL MODELS OF DEMAND AND SUPPLY
IN DIFFERENTIATED PRODUCTS INDUSTRIES

Amit Gandhi
Aviv Nevo

Working Paper 29257
<http://www.nber.org/papers/w29257>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2021

We thank Pierre Dubois, Phil Haile, Ali Hortaçsu, Felipe Kup Barbieri de Matos, Jing Tao and three referees for helpful comments and insightful suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Amit Gandhi and Aviv Nevo. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Empirical Models of Demand and Supply in Differentiated Products Industries
Amit Gandhi and Aviv Nevo
NBER Working Paper No. 29257
September 2021
JEL No. C01,D12,D22,D43,L13

ABSTRACT

This is an invited chapter for the forthcoming Volume 4 of the Handbook of Industrial Organization. We present empirical models of demand and supply in differentiated products industries with an emphasis on the key ideas arising from the recent applied literature. We start with a discussion of the challenges in modeling and estimation of demand for differentiated products, and focus on discrete choice characteristics-based demand models that address these challenges while allowing enough flexibility to capture realistic substitution patterns. Our discussion emphasizes how empirical strategies can leverage different features of data depending on the sources of variation that are commonly found in applied work. Moving to the supply-side, we show how demand estimates combined with a pricing model, can be used to recover markups and marginal costs. We also show how the model of pricing can be tested. We discuss a baseline Bertrand-Nash model of competitive pricing, and expand it to cover a) coordinated pricing, b) wholesale relationships, and c) bargaining. We end the chapter with extensions of the demand model, including dynamic and continuous demand.

Amit Gandhi
University of Pennsylvania
akgandhi@sas.upenn.edu

Aviv Nevo
Department of Economics
The Ronald O. Perelman Center
for Political Science and Economics
133 South 36th Street, Room 617
Philadelphia, PA 19104
and NBER
anevo@wharton.upenn.edu

Contents

1	Introduction	3
2	A Motivating Example	5
2.1	Model	7
2.2	Estimation and Results	11
2.3	Discussion	12
3	Demand	12
3.1	Background	13
3.2	Discrete Choice Demand Models	15
4	Demand Estimation	25
4.1	The Estimation Problem	25
4.2	What Variation in the Data Can Identify the Parameters?	27
4.3	The General Estimation Procedure	33
4.4	Extensions	45
5	Supply	49
5.1	The Workhorse Model of Horizontal Competition	50
5.2	Distinguishing Between Models of Competition	53
5.3	Adding Retailers Into the Mix	58
5.4	Models of Bargaining	61
6	Extensions of the Demand Model	67
6.1	Extensions to the Static Demand Model	67
6.2	Dynamic Demand	72
7	Concluding Comments	82

1 Introduction

To some, many cars might seem identical and all brands of cereal might seem as essentially the same. The typical consumers of these products beg to differ and seem to be willing to pay a premium to get the product they prefer. Indeed, most products are differentiated, at least to some degree. Empirical studies of these industries need to take this into account. However, modeling demand and supply in differentiated products is challenging. The last 25 years have seen significant modeling advances that have allowed industrial organization (IO) economists to make great strides in studying differentiated products industries. In this chapter we review some of the models that have allowed this progress.

The typical paper in this literature starts by writing down a model of demand. There are several reasons the literature has focused on demand. First, in order to answer many questions, for example the change in consumer welfare due to a merger, regulation or the introduction of new goods and services, we need an understanding of consumers' willingness-to-pay, and therefore the demand system. Second, and maybe most importantly, IO economists have used exogenous variation in demand conditions to estimate costs parameters, and at times, the model of competition. If the demand function is known then one can back out marginal costs implied by different supply models. Indeed, one can go further and identify the model of supply. Finally, armed with estimates of both demand and supply researchers can address not only a wide range of traditional IO questions, but also questions in other fields such as health, finance, taxation, housing and school choice, development, environmental policy and political economy and many others.

The core logic above is similar to methods used in homogeneous goods industries (Bresnahan, 1982, 1989). Indeed, the literature we review shares many characteristics with the literature surveyed by Bresnahan (1989). Like the earlier literature, the papers we survey will tend to focus on single industries and model the idiosyncrasies that make each industry unique. Variation will usually be across "markets", which are either defined by a time series (i.e., the same industry over time), a cross section (usually across different geographies), or panel data that combines both sources. Like the earlier literature, marginal costs are generally not going to be observed, and when needed they will be inferred or estimated.

There are however a few differences between the literature we review and the earlier literature. First, the newer literature will tend to focus on industries with differentiated

products. Indeed, much of the effort will be focused on modeling the differentiation in tractable ways. Second, for the most part there will not be an explicit “conduct parameter” to characterize the supply side that will be estimated. Instead, the literature will tend to focus on a small number of alternative supply models, in some case a single model.

Unlike some strands of the earlier literature, such as the “Structure-Conduct-Performance-Paradigm” (Bain, 1951) or the “New Empirical Industrial Organization” (Bresnahan, 1989), this literature does not have a distinct name. Some might say that it is a direct extension of the New Empirical IO, and therefore does not need a unique name, while others object to that characterization. Either way, it seems like the literature has gone nameless because it is viewed as synonymous with (modern) empirical IO. As such, the material we discuss below is a basis for much of the research done by the IO profession as well as the chapters that follow in this Handbook. We start our discussion in Section 2 with an example from Bresnahan (1987) that studies competition in a differentiated-products industry: the U.S. automobile industry. It might seem odd that we start the discussion of the modern literature with a paper that is almost 35 years old. We do so for several reasons. This paper provides a natural link between the recent and older literature. It also motivates and highlights several modeling challenges the modern literature has needed to address relative to the constraints imposed by the older literature. Finally, it offers an example of how modern IO combines demand and supply to answer a central question, some might say even the key question in IO, namely, the study of market power.

Having motivated why we care about demand estimation, we discuss in Sections 3 and 4 the canonical characteristics-based demand model and its estimation. We focus mostly on estimation using aggregate, market-level, data, but also discuss micro data (i.e., data where individual choices are observed) in order to gain intuition for the empirical problem. These sections complement the treatment in Berry and Haile (2021), who provide more conceptual backdrop to the model and deeper identification foundations of it. The material also overlaps with Akerberg et al. (2007) and Reiss and Wolak (2007) who offer a more econometric treatment of some of the earlier papers, and Dubé (2019) who offers a more theoretical discussion of demand models.

In Section 5 we turn from demand to supply and introduce the workhorse supply model of Nash-Bertrand price competition as well as various extensions of it. Our discussion is a natural segue to some of the other chapters in this Handbook (e.g., Lee et al. (2021)) that build on the modeling approach we discuss, extend it in various direction and apply it to various questions economic questions. In Section 6 we introduce several

extensions of the static differentiated product demand model presented earlier, including the generalization to dynamic demand.

2 A Motivating Example

In this section we provide a motivating example, based on Bresnahan (1987).¹ We have several goals. First, this example helps motivate why IO economists are interested in estimating demand. As we will show, knowing the demand function allows us to estimate markups and test models of competition. Second, we believe that this example provides a natural linkage to an earlier literature, reviewed in Bresnahan (1989), which studies some of the same economic questions as we do, but in homogeneous good industries. Finally, some of the basic ideas that are the foundation of more recent papers were laid out in this paper. At the same time, the paper highlights, and allows one to appreciate, some of the modeling contributions made by the more recent literature.

Bresnahan (1987) studies competition in the U.S. automobile industry in the mid 1950s. He notes that in 1955 more autos were sold, and prices were lower, relative to 1954 and 1956. He asks why this was the case. Specifically, he asks whether the prices observed in 1955 were the result of a “price war”, i.e., a breakdown in collusion in this industry.

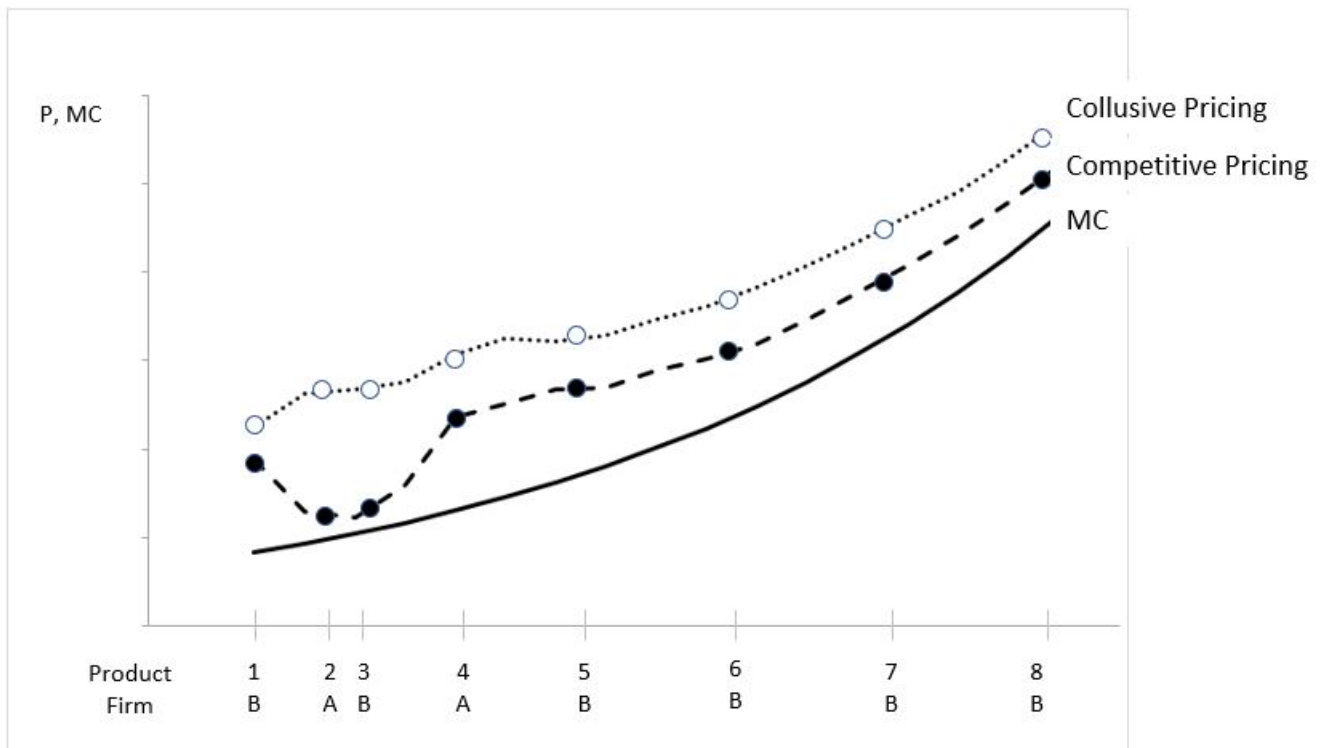
To infer the model of competition, he uses variation in demand conditions across different cars. The basic idea can be seen by examining Figure 2.1, which is a modified rendering of Figure 2 in Bresnahan (1987). Each point in the figure represents a product produced and priced by one of two firms, *A* and *B*. The vertical axis represents the price/cost, and the horizontal axis represents the quality of the product. The labels on the horizontal axis denote both the product number and which firm produces this product. The solid line represents the marginal cost. Finally, the two dashed lines depict the equilibrium prices under collusion and under non-collusive, i.e., competitive, pricing.

We note that marginal cost is increasing in quality and therefore under both pricing models, price also increases with quality. The products are differentiated, so in both cases the markups are positive. However, the markups are higher under collusive pricing. As we will see later, this can be helpful in distinguishing between collusive and competitive pricing in cases where we have a sense of what the markups are. The main feature of the graph, and what we will use to distinguish between the pricing models, is in how the markup differs with the proximity of competition. This is best seen by zooming in

¹See also Bresnahan (1981).

on products 2 and 3. Product 2 is priced by firm *A* and product 3 by firm *B*. In the competitive outcome their markups are low, because neither is very differentiated from the competition. However, in the collusive outcome their markups are closer to markups of other products because the proximity to a product priced by another firm is not putting any downward pressure on their own pricing. This suggests an experiment to distinguish competition and collusion: taking the location of products in characteristics space as given, if markups do not vary enough with proximity to competition then we might be able to reject a model of competition.

Figure 2.1: Intuition for Identification



Notes: This figure is a modified rendering of Figure 2 in Bresnahan (1987). Each point is a product. The vertical axis represents the price/cost, and the horizontal axis represents the quality of the product. The labels on the horizontal axis denote the product number and the firm that produces this product. The solid line represents the marginal cost. The dotted lines display the equilibrium prices under collusive and under competitive pricing.

While the conceptual experiment of distinguishing between competition and collusion in the above picture is clear, the key challenge is that several important constructs in

the graph are unobserved from data on prices and quantities. In particular, marginal cost, needed to construct the cost curve is not observed and therefore markups are unobserved. Note, that observing higher prices is not sufficient to separate competition and collusion, at least without further assumptions, since higher prices can be driven by higher costs or higher markups. We also do not know in what order to place the products on the horizontal axis, the distance between them, or for that matter whether a one-dimensional line adequately describes the spacing of products. This too, is crucial for the above conceptual experiment since it determines the closeness of competition amongst products.

To implement the conceptual experiment we need to estimate demand. Among other things, this allows us to measure the proximity of competition, and jointly with a pricing model to infer markups and marginal costs. This is what Bresnahan, and much of the literature discussed in this chapter, does. For reasons that we will explain in the next section, the approach that most of the literature has taken is to model a product as a bundle of pre-fixed observed characteristics that determine both demand and marginal cost. The parameters of the demand and cost functions will be identified from the variation in the distance between products in characteristics space, which Bresnahan assumes is exogenous, and how demand and pricing vary with this distance.

2.1 Model

We now discuss the specifics of the Bresnahan model to see how the conceptual ideas that come out of Figure 2.1 can be implemented and to illustrate some of the modeling issues and choices that need to be confronted.

2.1.1 Supply

Let $f = 1, \dots, F$ denote firms and $j = 1, \dots, J$ denote products operating in a single market t . Assume that each firm maximizes profits over some subset, \mathcal{J}_f , of the J products. Further assume that production costs are given by fixed costs, FC_j that vary by product, and marginal cost, mc_j , that vary by product as a function of its quality, but do not vary with quantity (i.e., there are no economies of scale).

Let p_j denote the price of product j , and bold face \mathbf{p} denotes the J dimensional (column) vector of all prices in this market. We will treat prices as endogenous (i.e., determined inside of the model), but the quality of the product as exogenous (i.e., determined

outside of the model). The profits of firm f are given by

$$\pi_f = \sum_{j \in \mathcal{J}_f} [(p_j - mc_j)q_j(\mathbf{p}) - FC_j],$$

where $q_j(\mathbf{p})$ is the quantity sold of product j , which is a function of the prices of all the J products.

Define an “ownership”, or conduct, structure as a J by J matrix, H , with elements equal to

$$H_{jk} = \begin{cases} 1, & \text{if } \exists f : \{j, k\} \subset \mathcal{J}_f; \\ 0, & \text{otherwise} \end{cases} \quad j, k = 1, \dots, J. \quad (2.1)$$

The elements of H equal to either 0 or 1. A value of 1 means that the two products, represented by the row and column indices, are priced as if jointly owned. This allows us to nest various pricing models. For example, pricing by single-product firms will have an identity matrix as the ownership matrix. At the other extreme, joint maximization of profits from all products will have a matrix of 1's.² Thus different models of firm behavior, such as whether firms compete or collude, map to different configurations of zeroes and ones in the ownership matrix H . Finally, let Ω be a J by J matrix with elements given by $\Omega_{jk} = -\partial q_k / \partial p_j \cdot H_{jk}$, where j indexes rows and k columns.

Using this notation we can write the first-order conditions of the firms' profit maximization problem as

$$\mathbf{q}(\mathbf{p}) - \Omega(\mathbf{p} - \mathbf{mc}) = 0,$$

where \mathbf{q} and \mathbf{mc} denote J -dimensional (column) vectors of the quantities and marginal costs. This in turn, implies a pricing equation

$$\mathbf{p} = \mathbf{mc} + \Omega^{-1}\mathbf{q}(\mathbf{p}). \quad (2.2)$$

Assuming the existence of a pure-strategy Nash-Bertrand equilibrium in prices and that the prices that support it are strictly positive, these first-order conditions characterize the equilibrium. If we know the demand derivatives, which enter Ω , we can use this relation,

²For now we do not consider values between 0 and 1. We return to this later in the chapter.

together with observed prices to compute implied cost and markups

$$mc = p - \Omega^{-1}q(p) \quad \text{and} \quad p - mc = \Omega^{-1}q(p). \quad (2.3)$$

In other words, for a given ownership structure, or model of competition, and using estimates of demand substitution, we are able to measure price-cost margins without observing cost data. Furthermore, we can compute these margins under different ownership structures, i.e., different H matrices, which, as we will see below, allows us to test different models of competition. This in essence formalizes the the conceptual experiments that we demonstrated in Figure 2.1.

As we will see in the rest of this chapter, this supply equation can be used in a variety of different ways.

2.1.2 Demand

The combination of a model of pricing and knowledge of the demand price derivatives allows Bresnahan to recover margins without knowing cost data.³ A key question is how to estimate the demand derivatives, or elasticities, given aggregate data on prices and quantities across products in a given year. Bresnahan uses a specific discrete-choice model, of vertical differentiation (Shaked and Sutton, 1983). In the next section we will discuss how this model can be generalized.

Let $i = 1, \dots, I$ denote consumers. A consumer i gets (indirect) utility from product j given by

$$\nu_i \cdot \text{quality}_j + y_i - p_j,$$

where quality_j is the product's quality, p_j is product j 's price, ν_i denotes the consumer's "taste" for quality, which can be viewed a willingness to pay for quality, and y_i is the consumer's income. In this model, all consumers evaluate a product's quality the same way, i.e., products are vertically differentiated. However, consumers differ in their willingness to pay for this quality and therefore consumers differ in the product they choose. Assume each consumer chooses exactly one of the J products or the outside option, of not purchasing a product (a new car in this application). If a consumer chooses the outside

³The idea of recovering unobserved marginal cost from the information in demand elasticities and the first-order conditions of the firm's optimal pricing behavior dates back to Rosse (1970) and Bresnahan (1981).

option, they get utility that is a function of the quality and price of the outside option, both captured by parameters that Bresnahan will estimate.

In this model, the only reason different consumers make different choices is because they have different ν_i 's. Therefore, to compute aggregate demand one needs to compute the set of ν 's that will induce a choice of each product and then integrate the mass in this region to get aggregate demand. In the vertical differentiation model, the sets are defined by cutoffs in ν . Namely, a consumer with a willingness to pay ν_i will choose product j if $\nu_{j+1}^* > \nu_i \geq \nu_j^*$, where products are ranked from the lowest quality ($j = 1$) to the highest ($j = J + 1$) and the cutoff ν_j^* is defined as the ν of the consumer who is indifferent between option j and option $j - 1$.⁴ This implies that the demand for product $j = 1, \dots, J + 1$ is given by

$$q_j = I[F(\nu_{j+1}^*) - F(\nu_j^*)],$$

where I is the number of consumers, and $F(\cdot)$ is the cumulative distribution function of ν . Bresnahan assumes a uniform distribution $U[0, V_{max}]$ in which case the expression for the cutoffs, aggregate demand and the own- and cross-price derivatives of demand have simple closed-form solutions. For example, the cross-price derivatives of demand are given by

$$\frac{\partial q_j}{\partial p_r} = \begin{cases} I \left[\frac{1}{\text{quality}_j - \text{quality}_r} \right] & r = j - 1, j + 1 \\ 0 & \text{otherwise} \end{cases}.$$

The price derivatives illustrate the restrictiveness of the demand model. Competition is highly localized. Each product only directly substitutes to at most two products: the product just above in quality space and the one just below (assuming these exist). This is a very strong assumption, which is driven by the scalar restriction that products can be placed along a one-dimensional quality measure.

For estimation, Bresnahan assumes that the quality of product j is a function of K characteristics, $x_j^{(k)}$, $k = 1, \dots, K$, observed by the firms, consumers and importantly by the researcher. He assumes that quality is given by $\sqrt{\beta_0 + \sum_k \beta^{(k)} x_j^{(k)}}$, where $(\beta_0, \dots, \beta_K)$ are parameters to be estimated.

There are several assumptions baked into this setup. The obvious one is the functional form of quality. More importantly, there is no heterogeneity in the coefficients, namely in

⁴For completeness $\nu_0^* = 0$ and $\nu_{J+1}^* = V_{max}$, namely the upper bound of the distribution of ν , or infinity if the distribution is unbounded.

how consumers value the characteristics, and all the relevant characteristics are observed by the researcher. Relaxing these constraints is a central focus of the literature that followed.

2.2 Estimation and Results

Bresnahan estimates the model using annual U.S. list prices and quantity produced by name plate. He abstracts away from manufacturer-dealer relations, negotiations in setting prices and price dispersion within the year and across geography and consumers, and further assumes that all cars produced are sold that year in the U.S. market. He also uses information on the characteristics of cars.

He assumes that marginal cost is a parametric function of observed characteristics.⁵ Finally, he assumes that the observed prices and quantities, $\{p_j, q_j\}_{j=1}^J$ are given by

$$p_j = p_j^* + \epsilon_j^p \text{ and } q_j = q_j^* + \epsilon_j^q, \quad (2.4)$$

where $p^*(\mathbf{x}; H, \theta)$ and $q^*(\mathbf{x}; H, \theta)$ are the equilibrium prices and quantities predicted by the model, \mathbf{x} is a $(J \times K)$ matrix of the characteristics of all products, θ is a vector denoting the parameters of the model, ϵ_j^p and ϵ_j^q are i.i.d. zero mean normally distributed shocks. Note that these errors are not part of the model and are best viewed as errors in the measurement of prices and quantities. We will refer to this way of setting up the econometric error terms, as non-structural, and sometimes refer to the error terms as “add-on” errors. This is an area where the more recent research took a different approach.

He estimates, separately for each year, four different models using maximum likelihood. The first three models are variants of the model described above where the ownership matrix takes on three sets of values: (i) joint ownership (he refers to this model as collusion); (ii) current ownership (he refers to this as Nash); and (iii) single product ownership. He also estimates a model non-nested in the above where $p_j^* = \exp[\alpha_0 + \sum_k \alpha_k x_j^{(k)}]$ and $q_j^* = \exp[\lambda_0 + \lambda_1(P_j - P_j^*)]$.

He selects among the models in two ways. First, he uses a Cox test of non-nested alternatives.⁶ The results of this test (presented in Table 3 of the paper) reject all but the collusive model in 1954 and 1956, but in 1955 only the Nash model is not rejected.

⁵Specifically, he assumes that that $mc_j = \mu e^{quality_j}$, where μ is a parameter to be estimated. Note, that he does not allow for any unobserved factors to impact marginal cost.

⁶The likelihood ratio of the null and the alternative is the central statistic in this test. The mean and variance are computed under the null and used to compute a test statistic that is distributed as a standard

Second, he uses an informal test that compares estimates across years under different models. This informal test confirms the results of the Cox test. If we use the collusive model in 1954 and 1956 and Nash model in 1955 as the maintained assumption, then the structural parameters are generally steady and robust (Table 4 in the paper). However, if we keep the same model throughout the three years then the structural parameters vary between 1955 and 1954/56 (Table 5 in the paper). In other words, we can explain the change in 1955 in two ways: either the model of competition changed and the structural parameters were generally unchanged or there was a break in 1955, relative to 1954 and 1956, in preferences and cost. The latter does not seem very realistic and therefore the change in the model of competition seems like the reasonable explanation.

2.3 Discussion

Bresnahan (1987) offers a powerful method to infer demand and cost parameters together with the model of (price) competition. The basic idea utilizes the intuition we saw in Figure 2.1: taking the location of products in characteristics space as given, we can infer the model of competition by seeing how markups vary as a function of the distance to other products. Different models of price competition will predict different markups. We can distinguish between different models of competition by matching patterns in the data (as in the informal test), or by asking which model "better fits" the data.

As we pointed out, the specific demand model used by Bresnahan is quite restrictive. Competition is localized and is only between the immediate neighbors on the quality line. There is limited heterogeneity in preference and all product characteristics are assumed to be observed. Finally, the estimation is based on non-structural error terms, which some view as the main limitation of this approach. As we will discuss in the rest of this chapter, more recent work has built on the key insights above and relaxed several restrictions by considering more flexible functional forms and being explicit about the structural errors and the challenges they create.

3 Demand

As we saw in the motivating example in the previous section, demand plays a key role in the study of supply: it can be used (jointly with a pricing equation) to recover unob-

normal. The test requires that either the null or the alternative be true. For alternative tests of non-nested models see Vuong (1989) and Rivers and Vuong (2002).

served marginal costs and markups or to test different models of supply. This has led to a significant IO literature focused on demand estimation. In this section we focus on static models of demand for differentiated products proposed in the literature. We start with a discussion of the difficulties in estimating demand for differentiated products. We next discuss the various solutions offered in the literature, with a focus on discrete choice models.

3.1 Background

The empirical analysis of consumer demand has a long and rich history in economics and econometrics.⁷ Since Stone (1954) researchers estimating demand systems have tried to balance flexible functional forms and a connection to economic theory. Examples include the Rotterdam model (Theil, 1965; Barten, 1966), the Translog model (Christensen et al., 1975), and the Almost Ideal Demand System (Deaton and Muellbauer, 1980). Deaton (1986) offers a comprehensive review of this literature. These models cannot directly be applied to estimating demand for differentiated products. To understand why consider the following.

Suppose we want to estimate demand for J differentiated products in market t . In principle, the most straight-forward approach is to write down an aggregate demand system of the form

$$q_{jt} = Q_j(\mathbf{p}_t, \mathbf{x}_t, \boldsymbol{\xi}_t), \quad j = 1, \dots, J, \quad (3.1)$$

where q_{jt} is the quantity demanded of product j in market t , $Q_j(\cdot)$ is the demand function for product j , \mathbf{p}_t is a $J \times 1$ vector of prices, \mathbf{x}_t is a $J \times K$ matrix of (observed) variables that shift demand, and $\boldsymbol{\xi}_t$ is a $J \times 1$ vector of unobserved demand shocks. Note, that in general quantity demanded of each product is a function of prices, observed variables and demand shocks of all products. This approach, while intuitive, ends up being problematic when modeling demand for differentiated products.

First, as the number of options, J , becomes large there is a dimensionality problem due to the large number of parameters to be estimated. For example, consider a simple linear demand system,

$$\mathbf{q}_t = A\mathbf{p}_t + \boldsymbol{\epsilon}(\boldsymbol{\xi}_t) \quad (3.2)$$

⁷See Schultz (1938) and Stigler (1954) for surveys of the very early work.

where \mathbf{q}_t is a $J \times 1$ vector of quantities, A is $J \times J$ matrix of parameters and $\epsilon(\boldsymbol{\xi})$ is a vector of econometric error terms, which are a function of the unobserved demand shocks in equation (3.1). Note, that this stylized system is restrictive in several ways: prices enter linearly, we omitted the dependence on the observable variables, \mathbf{x}_t , and imposed a strong restriction of how the demand shocks, $\boldsymbol{\xi}_t$ enter the model (see Berry and Haile (2021) for a discussion of this point). Even with these restrictions, this system implies J^2 parameters to be estimated. The number of parameters to be estimated can be somewhat reduced by imposing symmetry of the Slutsky matrix and other constraints implied by economic theory, but the number of parameters to be estimated is still proportional to J^2 , and too large to be manageable for a large number of products. Of course, with a more flexible functional form, the problem becomes worse.

Second, in some cases the key object of interest is not aggregate demand, but a model of individual consumer choice: for some applications we would like to explicitly model and estimate the distribution of heterogeneity. The above approach, generally, does not let us do this.

Third, this demand system does not easily allow us to predict the demand for new goods. Once we relate products to their characteristics we would be able, to some degree, to predict the demand for new goods. How well we can predict the demand depends on the importance of unobserved demand shocks.

Finally, estimating the above demand system usually faces several empirical problems. Prices of narrowly defined products typically are highly collinear, making it difficult to separately identify the price effects of individual products. This problem is augmented once we have many prices on the right hand side, since we typically think that prices are correlated with the error terms and therefore the numbers of required instrumental variables (IVs) increases. Finding a single IV is not easy, making it almost impossible to find enough IVs that are both exogenous and will not generate moment conditions that are not nearly collinear.

Several approaches have been proposed to deal with these issues, largely by micro-founding the preference structure that underlies demand. For example, a popular approach in the trade literature, which is helpful with the dimensionality problem, is to impose symmetry across products in a representative agent's preferences over products.

A leading example of a model that imposes strong symmetry assumptions is the constant elasticity of substitution (CES) demand model (Spence, 1976; Dixit and Stiglitz, 1977).⁸

Another approach that is somewhat more popular in IO, yet still relies on demand systems in product space, is the multi-level demand system proposed by Hausman et al. (1994) and Hausman (1996). The model builds on ideas from the multi-stage budgeting literature (see Deaton and Muellbauer (1980) for a discussion) to construct a multi-level demand system for differentiated products. The typical implementation has three levels: demand for an overall category (say breakfast cereal), demand for segments within the category, taking category demand as given, and demand for brands within a segment, taking segment demand as given. Each level allows for a flexible functional form. This approach can somewhat help with the dimensionality problem but still suffers from the other issues discussed above.

3.2 Discrete Choice Demand Models

The approach most commonly used in IO for estimating demand for differentiated products, and the focus of this chapter, views a product as a collection of characteristics rather than qualitatively different products (Gorman, 1956; Lancaster, 1966; Rosen, 1974). The basic idea is somewhat similar to what we saw in Section 2: substitution between products will be driven by their characteristics. Products that are similar in their characteristics will be closer substitutes. To see how this helps with the dimensionality problem, we can reconsider the linear demand system in (3.2). The matrix A will be a function of product characteristics, and parameters, and therefore the relevant dimension is the number of the characteristics, and not the number of products. The model also offers a natural way to include additional (unobserved) characteristics that impact demand as well as demand shocks more generally.

The specification of the model starts with a random utility, which is a function of observed and unobserved (by the researcher) product characteristics, including prices. We focus on a linear utility model and assume that the (conditional indirect) utility of

⁸This approach is rarely used in IO for the reasons discussed in Nevo (2011). Interestingly, this is despite the similarities between the CES model and the Logit model (Anderson et al., 1992; Dubé et al., 2021), which is heavily used and discussed below.

consumer $i, i = 1, \dots, I_t$ in market $t, t = 1, \dots, T$, from product $j, j = 1, \dots, J^9$ is given by

$$u_{ijt} = x_{jt}\beta_{it} + \alpha_{it}p_{jt} + \xi_{jt} + \varepsilon_{ijt}, \quad (3.3)$$

where $x_{jt} \in \mathbb{R}^K$ is a (row) vector of observed product characteristics, $p_{jt} \in \mathbb{R}$ is the price of product j in market t and $\xi_{jt} \in \mathbb{R}$ is a demand shock that is observed by consumers and firms, but not by the researcher. As before, bold face, $\mathbf{x}_t, \mathbf{p}_t$ and $\boldsymbol{\xi}_t$ will denote the collection of x_{jt}, p_{jt} and ξ_{jt} across j within a market t . Finally, the model includes an idiosyncratic taste shock, ε_{ijt} , which captures randomness in choices: a consumer faced with the same choice set (and prices) might make different choices at different times. ε is typically assumed to be i.i.d. across (i, j, t) and most often specified as a draw from a type-1 extreme value distribution (with a scale parameter normalized to 1), yielding the Mixed Logit model (sometime referred to as "random coefficients Logit model").

Note that the utility from a product only depends on its own characteristics (and prices). If the utility were to depend on the characteristics of all products then we would be back to the dimensionality problem discussed in the previous subsection. Individual choices, and therefore aggregate demand, will depend on the relative utility from products, and therefore the characteristics of all products (as in equation (3.1)). However, restricting the way the characteristics enter utility (in a way that seems quite natural) will allow us to write a model that is both consistent with equation (3.1) and reasonable to estimate even with a large number of products.

An important part of this specification, and what distinguishes it from much of the earlier discrete choice literature, is the unobserved characteristic, ξ_{jt} . This characteristic captures unobserved characteristics of the product, factors that are difficult to quantify, brand equity, systematic shocks to demand, or unobserved promotional activity. In working with market-level data this unobserved characteristic is essential to explain cross market variation: as Berry et al. (1995) noted, without ξ_{jt} , if we compare actual to predicted demand shares given the large number of consumers in a usual market, we will be left rejecting the demand specification. The unobserved characteristic ξ_{jt} helps rationalize the wedge between actual and predicted demand.

At this point it might not be clear why we need to separate ξ_{jt} from ε_{ijt} : mathematically, ξ_{jt} is only shifting the mean of ε_{ijt} , by j and t . However, we will assume that $\boldsymbol{\xi}_t$ is observed by firms before setting market-level prices, while the individual realizations of

⁹To simplify notation, we assume each market has the same number of products J does not depend on t .

ε_{ijt} do not impact pricing. When estimating the model using market-level data, ξ_{jt} will typically end up being the econometric error term and therefore prices, as well as other choice variables, could be correlated with it.

From a modeling prospective, it is also important to recognize that ξ_{jt} does not vary within a market. Empirically, one needs to define both the geographical and temporal boundaries of a "market." For example, is a market defined as a city, state, nation, or the world? Is each day, week, month or year a different market? The answers are application-specific and need to account for institutional detail as well as data considerations.

The parameters α_{it} and β_{it} capture the relative weight that consumers put on price and product characteristics. Let $\beta_{it}^{(k)}$ denote the weight consumer i puts on characteristic $x_{jt}^{(k)}$. It is typically modeled as

$$\beta_{it}^{(k)} = \beta_0^{(k)} + \sum_{l=1}^L \beta_d^{(l,k)} D_{ilt} + \beta_\nu^{(k)} \nu_{it}^{(k)}. \quad (3.4)$$

Observe there are 3 elements composing a consumer i 's taste for a characteristic k . The parameter $\beta_0^{(k)}$ is common to all consumers. Heterogeneity in consumers' taste around the common taste components is modeled as a function of a set of L "demographic" variables (e.g. income, age, or family size), $D_{it} = (D_{i1t}, \dots, D_{iLt})^\top$, as well as a random variable $\nu_{it}^{(k)}$. The differences between the demographic and random variables is that the demographic variables are assumed to either be observed, or that their distribution is known or can be estimated.

The usefulness of the above formulation is that it allows, in principle, to capture different forms of heterogeneity. For example, a younger consumer might like cereal with more sugar, while an older consumer might either not like sugar or have a weaker preference for it. The random components, $\nu_{it}^{(k)}$ capture variation in preferences above and beyond what standard demographics can explain.

The price coefficient, α_{it} is modeled in a similar way

$$\alpha_{it} = \alpha_0 + \sum_{l=1}^L \alpha_l D_{ilt} + \alpha_\nu \nu_{it}^{(0)}. \quad (3.5)$$

In some cases researchers might specify the price coefficient using logs to ensure that the distribution of price coefficients is negative for all consumers.

The specification of the demand system is completed with the introduction of an outside good: consumers may decide not to purchase any of the J inside products. The indirect utility from this outside option, indexed as $j = 0$ is

$$u_{i0t} = \varepsilon_{i0t},$$

where the non-idiosyncratic part of utility from the outside good is normalized to zero. Note, that the non-idiosyncratic part of utility for the inside goods should be interpreted as being the incremental utility relative to the outside good.

In principle, the specification of the (conditional indirect) utility in equation (3.3) should include not just price, p_{jt} , but rather $y_i - p_{jt}$, where y_i is income. Because of the (quasi) linear specification this has no impact: income enters linearly into utilities from all options, including the outside good, therefore it will not impact choice probabilities since only the difference in utilities matter for choice probabilities. In order to simplify the exposition we dropped income out of equation (3.3). However, if $y_i - p_{jt}$ enters utility non-linearly, or interacted with other variables, income will not cancel and should be explicitly included.

For what follows it is useful to define

$$\delta_{jt} = x_{jt}\beta_0 + \alpha_0 p_{jt} + \xi_{jt} \quad (3.6)$$

as the “mean utility” for product j in market t . Let Γ be a $(K + 1) \times L$ matrix with the coefficients of the demographic variables in equations (3.5) and (3.4), Σ be a $(K + 1) \times (K + 1)$ diagonal matrix with the diagonal equal to $(\alpha_v, \beta_v^{(1)}, \dots, \beta_v^{(K)})$, and $\nu_{it} = (\nu_{it}^{(0)}, \dots, \nu_{it}^{(K)})^\top$. The consumer-level variation across this mean utility is captured by two terms. The first, $\mu_{ijt} = (x_{jt}, p_{jt}) \cdot (\Gamma D_{it} + \Sigma \nu_{it})$, captures the interaction of consumer taste preferences and product characteristics. The second, is the random term, ε_{ijt} . Before we make any distributional assumptions the two terms are interchangeable.

We assume that consumers choose a single option that gives the highest utility.¹⁰ This allows us to derive purchase probabilities by integrating over the distribution of ε , and market shares by integrating over the mixing distribution. For example, for Mixed Logit and the specification in equations (3.4) and (3.5) the market shares are given by

$$s_{jt} = \sigma_j(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t; \Gamma, \Sigma) = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k=1}^J \exp(\delta_{kt} + \mu_{ikt})} dF(D_{it}, \nu_{it}). \quad (3.7)$$

¹⁰In section 6 we discuss how we can relax this assumption and deal with situations when consumers purchase multiple brand or multiple units of the same brand.

where $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})$.

3.2.1 Price Elasticity and Substitution Patterns

To compute the market shares given by equation (3.7) we need the mixing distribution, both the distribution function, F , and the parameters (Γ, Σ) . The key to recognize is that different mixing distributions will imply different patterns of substitution among products.

Possibly the simplest assumptions we can make is to eliminate the interactions between the consumer attributes and the product characteristics (either by setting $\Gamma = 0$ and $\Sigma = 0$, or by assuming that the distribution $F(D_{it}, \nu_{it})$ is degenerate). This restriction yields the Logit model and the market share of brand j in market t , is given by

$$s_{jt} = \frac{\exp(\delta_{jt})}{1 + \sum_{k=1}^J \exp(\delta_{kt})}. \quad (3.8)$$

The Logit model is very tractable and can be estimated using linear methods. However, it significantly restricts substitution patterns. At a high level, it is the opposite of the vertical differentiation model we discussed in Section 2. In the vertical differentiation model competition is localized to just the immediate "neighbors." In the Logit model competition is global and depends only on the market share, and not how close the products are in characteristics space.¹¹ To see this consider the price elasticities implied by the Logit model

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} \alpha_0 p_{jt} (1 - s_{jt}) & \text{if } j = k \\ -\alpha_0 p_{kt} s_{kt} & \text{otherwise} \end{cases}.$$

There are two patterns that emerge from these elasticities. First, consider the own-price elasticities. Since we have many products, generally, the market share of any given product is small, and therefore $\alpha_0(1 - s_{jt})$ is nearly constant. Therefore, the own-price elasticities are proportional to price: the lower the price, the lower the elasticity (in absolute value). When these elasticities are used with the pricing model presented in Section 2.1.1 they predict a higher markup for the lower-priced products. This is a somewhat surprising pattern, which nevertheless might be correct in some industries. The key is not whether this pattern is correct or not but that it is driven completely by modeling

¹¹In this sense the Logit model is similar to the demand models that impose symmetry. Indeed, as Anderson et al. (1992) show, the Logit model can be formally represented with a representative agent utility that is somewhat similar to the CES.

assumptions and not informed in any meaningful way by the data. In other words, empirically finding such a pattern using the Logit model is not a “finding” but rather a direct implication of the modeling assumptions.

Second, consider an increase in the price of product k . We would generally expect that consumers, who decide to no longer purchase the product because of the price increase, will substitute to similar products. For example, if the price of a BMW sedan increases we would expect consumers to substitute more to other luxury sedans than to, say, a Honda Civic. In the Logit model this is not the case. The modeling assumptions imply that the substitution to product j when the price of k increases is given by $\frac{\partial s_{jt}}{\partial p_{kt}} = \alpha_0 s_{kt} s_{jt}$. The fraction of consumers who leave product k and switch to product j , also known as the diversion ratio, is given by $\frac{\partial s_{jt}}{\partial p_{kt}} / \frac{\partial s_{kt}}{\partial p_{kt}} = s_{jt} / (1 - s_{kt})$. In other words, the Logit modeling assumptions imply that substitution, and diversion, is proportional to market share and not to how close the products are. As before, this might be (approximately) correct in some industries. The key, however, is that this pattern is totally driven by a modeling assumption and is not informed by the data.

This property of the Logit model is closely related to the so-called independence of irrelevant alternative (IIA) property of Logit: the relative probability of choosing product k or product j does not depend on the existence (or characteristics) of other alternatives. A similar property holds in the aggregate, namely that the relative market share s_{jt} / s_{kt} does not depend on the characteristics of other products. The behavior of individual choice probabilities and market share are often confused as being the same, but they are not. Once we allow for heterogeneity in consumer tastes, the IIA property could hold at the individual level, but the aggregate property might not. This is the central value of the mixing distribution F in the model - to allow for more flexible substitution patterns in aggregate demand.

Why does the Logit model yield these predictions? Basically, it is the fact that the only heterogeneity in the model are the i.i.d. ε_{ijt} 's. So when the price of k increases, the consumers who no longer choose k will choose the other options at the same frequency as the “average” consumer, namely, in proportion to the market share. In reality, we think that consumers who no longer choose product k are more likely than the average consumer to choose similar options, as in the BMW example above. Another way of saying this, in a model with heterogeneity in preferences the consumers who choose a product k are selected and reveal something about their preferences. The i.i.d. assumption, implicit in the Logit model, shuts off this selection effect.

In order to capture richer substitution patterns we need to relax the i.i.d. assumption. The variation around the mean utility has to be correlated across options: a consumer who is more likely than average to buy a BMW should also be more likely than average to buy a similar car. This can be achieved in one of two ways in the model. First, we could generate the correlation by relaxing the i.i.d. assumption and allowing ε_{ijt} to be correlated across j .¹² Alternatively, we could generate the correlation by allowing for heterogeneity in tastes.

The Nested Logit model is an example of the first approach. As in the Logit model, we continue to assume that $\mu_{ijt} = 0$, but now we divide the products into mutually exclusive nests, or segments, $g = 1, \dots, G$. Finally, let $\varepsilon_{ijt} = \lambda \varepsilon_{ig(j)t} + \varepsilon_{ijt}^1$, where ε_{ijt}^1 is an i.i.d. extreme value shock, $\varepsilon_{ig(j)t}$ is a shock common to all options in segment g , and λ is a parameter that captures the relative importance of the two. Assuming a particular distribution for $\varepsilon_{ig(j)t}$ we get the Nested Logit model (Cardell, 1997). Note, that if $\lambda = 0$ we are back to the Logit model. The Nested Logit model is a special case of the more general Generalized Extreme Value model (McFadden, 1978, 1981), which imposes correlation among the options through correlation in ε_{ijt} .

A different solution to the problem with the elasticities is offered by the Mixed Logit or random coefficients Logit, as described by equation (3.3).¹³ An early version of this model was introduced by Boyd and Mellman (1980) and Cardell and Dunbar (1980), but popularity today was triggered following Berry et al. (1995) and McFadden and Train (2000). This model addresses both of the concerns with the elasticities by allowing for heterogeneity in preferences, which generates correlation in utility among products through μ_{ijt} . Thus, the heterogeneity in tastes for the product characteristics drives correlation in utility over products.

In this model, assuming the distribution of heterogeneity is given by equations (3.4) and (3.5), the price elasticities are

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_{it} s_{ijt} (1 - s_{ijt}) dF(D_{it}, \nu_{it}) & \text{if } j = k \\ \frac{p_{kt}}{s_{jt}} \int \alpha_{it} s_{ijt} s_{ikt} dF(D_{it}, \nu_{it}) & \text{otherwise} \end{cases} \quad (3.9)$$

¹²In principle one could consider estimating an unrestricted variance matrix of the shock, ε_{ijt} . This, however, reintroduces the dimensionality problem discussed above, since it involves estimating a number of parameters proportional to J^2 .

¹³Note, an alternative view of the Nested Logit model is to include in x_{jt} a nest dummy variable. By defining the distribution of ν_i appropriately (yet leaving ε_{ijt} i.i.d.) we are back to the Nested Logit, but now the correlation is motivated through the interaction of a product characteristic, the nest dummy, and heterogeneity in preference for this characteristic.

where s_{ijt} is the probability that consumer i purchases product j in market t . Now, each consumer has a different price sensitivity, which will be averaged to a product-specific mean price sensitivity using the individual probabilities of purchase as weights, and therefore the price sensitivity will be different for different products. So if, for example, product j has lower prices and is more likely to be purchased by price sensitive consumers, its average price sensitivity will be higher because the price sensitive consumers will receive higher weights. Therefore, own price elasticities are not driven solely by functional form, but by the heterogeneity in the price sensitivity across consumers who purchase the various products.

The Mixed Logit demand model allows for flexible cross-product substitution patterns, which are not constrained by a priori segmentation of the market (yet at the same time can take advantage of this segmentation by including a segment dummy variable as a product characteristic). In particular, as can be seen in (3.9), the correlation between μ_{ijt} and μ_{ikt} will induce correlation between s_{ijt} and s_{ikt} , and the latter correlation determines substitution patterns.

The modeling advantages of the full model do not come without a cost. It is significantly more complex to estimate. Furthermore the key in achieving all of these benefits is being able to estimate a meaningful degree of heterogeneity. We discuss these costs and empirical strategies for approaching them in the estimation section below.

3.2.2 Consumer Welfare

A common application of demand models is to compute welfare gains. This could be the main focus of the analysis or a side computation. For example, Trajtenberg (1989) and Petrin (2002) compute the welfare gains from the introduction of new goods. Nevo (2000a) computes the welfare implications of regulatory intervention, a merger in his case, and Pakes et al. (1993) compute a price index. The model discussed above can be used to compute welfare gains in these cases, by relying on the so-called inclusive value.¹⁴

McFadden (1978) defines the inclusive value (or social surplus) as the expected utility prior to observing $(\varepsilon_{i0t}, \dots, \varepsilon_{iJt})$. The expectation needs to account for a selection problem: the choice maximizes the utility given in equation (3.3) after observing $(\varepsilon_{i0t}, \dots, \varepsilon_{iJt})$. Therefore we need to compute the expected value of utility conditional on selection. When the idiosyncratic shocks ε_{ijt} are distributed i.i.d. extreme value, the inclusive value from a

¹⁴For non-parametric methods for welfare analysis of economic changes in setting of multinomial choice see Bhattacharya (2018).

subset $A \subseteq \{1, 2, \dots, J\}$ of the choice alternatives is defined as

$$\omega_{iAt} = \ln \left(\sum_{j \in A} \exp \{ \delta_{jt} + \mu_{ijt} \} \right). \quad (3.10)$$

Without heterogeneity the inclusive value captures the average utility in the population, up to a constant, averaging over the individual draws of ε , hence the term social surplus. When the utility is linear in price, or more precisely income minus price, the inclusive value can be converted into a monetary equivalent by dividing by the price coefficient. See McFadden (1981) and Small and Rosen (1981) for further details.

There are two somewhat distinct cases when we typically want to compute welfare. In the first case we observe a series of quantities and prices and we want to summarize them into a welfare measure. Nevo (2003) studies precisely this problem. A key issue that he points out is that the normalization of the utility from the outside good to zero, which is innocent for the purpose of estimating choice probabilities, is not purely a normalization when we want to compute a price index over time. The issue is not that the utility from the outside good is set to zero but that it is assumed to be constant over time. For example, suppose that in the data we see the share of the inside products going up over time. This could be because the price of the inside products decreased (or their quality increased) or because the outside option got worse. These have opposing welfare implications. Assuming that the outside good is normalized to zero rules out the latter.

The second case is one where we use the model to compute a welfare gain from a counterfactual outcome. Petrin (2002) is an example of such a case. He estimates the welfare gains from the introduction of minivans. To do so he creates a counterfactual outcome of what the equilibrium would have looked like without minivans. He then essentially reverts to the first case and uses the model to summarize the welfare effects from the (observed and simulated) price and quantities.¹⁵

The value of using the more flexible Mixed Logit model for welfare calculations differs somewhat between these cases. In the first case, both the Logit model and the more flexible Mixed Logit model will fit the market share data, as long as ξ_{jt} is allowed to vary by market (including over time). One can show that the welfare measure will equal $\ln(1/s_{0t})$ in the Logit model and $\int \ln(1/s_{i0t}) dF(D_{it}, \nu_{it})$ for the Mixed Logit model, where s_{0t} is the market share of the outside good and s_{i0t} is the probability of consumer i choosing the

¹⁵In principle, he could have used data from pre-introduction to conduct this analysis. This risks confounding the value of the introduction with other trends that are happening at the time.

outside option. The Logit model will yield a different answer as long as there is heterogeneity in the probability of choosing the outside good. In many cases we care about the change in welfare from period t to period $t - 1$, which is given by the difference between

$$\ln\left(\frac{1}{s_{0t}}\right) - \ln\left(\frac{1}{s_{0t-1}}\right) \quad \text{and} \quad \int \ln\left(\frac{1}{s_{i0t}}\right) dF(D_{it}, \nu_{it}) - \int \ln\left(\frac{1}{s_{i0t-1}}\right) dF(D_{it-1}, \nu_{it-1}).$$

Since both models perfectly fit the market shares, i.e., $s_{0t} = \int s_{i0t} dF(D_{it}, \nu_{it})$, the difference depends on the change in the heterogeneity in the probability of choosing the outside option, s_{i0t} . It is important to note that this difference can be positive or negative.

Things are a bit different in the second case. A common claim is that in this case the Logit model overestimates consumer gains. For example, in the case of new product introduction, the logic is that every new option introduced in the Logit model will mechanically increase welfare because it gives the consumer another draw from the distribution of ε . Since the chosen product is the option with the highest utility, the consumer's utility should increase with the availability of another option. The claim is that introducing heterogeneity decreases the "reliance" on the Logit error term and therefore diminishes this effect. Petrin (2002) argues the point empirically: when he introduces heterogeneity the computed welfare effects from the introduction of the minivans decrease. To understand what is driving these results we note that the exercise has two steps: generating a counterfactual and then summarizing the counterfactual (and observed) prices and quantities into a welfare measure. The first step is what largely generates the problem. The Logit model does a poor job of predicting the counterfactual equilibrium. This should not be surprising since we know that the Logit model does not do well at predicting marginal changes (i.e., substitution) and therefore it should not be surprising that it fails to do well in predicting non-marginal changes.

When we observe shares pre and post introduction (and ξ_{jt} can be different pre and post product introduction), the Logit model can match the data. In this case any "bias" introduced by the first step, of generating the counterfactual, is eliminated. The Logit model might still not give the "correct" welfare measure, but that is due to problem discussed in the first case above where heterogeneity in the demand for the outside good is ignored under the Logit model.¹⁶ This is confirmed by Berry and Pakes (2007) who offer a model where ε_{ijt} is dropped from equation (3.3), in a model they call the pure char-

¹⁶Nevo (2011) demonstrates this with the use of a classic example due to Debreu (1960) often called the "red-bus blue-bus example". He shows that the Logit model fails miserably in the first step of the analysis, in generating counterfactual market shares, in that example. But if we could eliminate the first

acteristics demand model.¹⁷ In principle, one would think that this would eliminate the concerns raised by Petrin (2002) about ε_{ijt} driving the welfare results. Instead, in their Monte Carlo results they find that using the pure characteristics model matters for the estimated elasticities (and mean utilities) but not the welfare numbers. They conclude that, consistent with the results in Nevo (2003, 2011), “the fact that the contraction fits the shares exactly means that the extra gain from the logit errors is offset by lower δ ’s, and this roughly counteracts the problems generated for welfare measurement by the model with tastes for products.” Akerberg and Rysman (2005) propose to use historical data to estimate how δ should change as a function of the number of products, in order to capture the change needed to rationalize the post-introduction data (if it were observed).

4 Demand Estimation

We have seen above the important role played by heterogeneity in consumer preferences for generating realistic patterns of product substitution and price elasticities. This flexibility comes at a cost: the more general model is more difficult to estimate. In this section we discuss the empirical problem of identifying and estimating the parameters of the market demand system with heterogeneous consumer preferences.

4.1 The Estimation Problem

The parameters to be estimated are defined in equation (3.7). They include (α_0, β_0) which are the parameters in the mean utility defined by equation (3.6). For reasons that will become clear below, we will refer to these as the “linear parameters.” Next, are the coefficients of the demographic variables in equations (3.5) and (3.4) captured by the matrix Γ of dimension $(K + 1) \times L$. Finally, the matrix Σ , a $(K + 1) \times (K + 1)$ diagonal matrix with the diagonal equal to $(\alpha_v, \beta_v^{(1)}, \dots, \beta_v^{(K)})$, captures the idiosyncratic “taste for characteristics.” Jointly, Γ and Σ will often be called the “non-linear parameters.” The full parameter vector to be estimated is $\theta = (\alpha_0, \beta_0, \Gamma, \Sigma)$.

In estimation it is typical to treat the distribution of the idiosyncratic “taste for characteristics” $\nu_{it} = (\nu_{it}^{(0)}, \dots, \nu_{it}^{(K)})$ as independent of the distribution of demographics D_{it} . In step, of predicting the counterfactual market by, say, observing the market shares post introduction, the Logit model can get the correct answer in that example.

¹⁷As we explained above, ε_{ijt} help rationalize observed choices. Indeed, once we drop them the model can in principle have difficulty rationalizing certain patterns of behavior. See Athey and Imbens (2007) for a discussion of the potential problems with the pure characteristics model and an alternative model.

other words,

$$F(D_{it}, \nu_{it}) = F_D(D_{it})F_\nu(\nu_{it}).$$

A further restriction often used in practice is to treat each $\nu_{it}^{(k)}$ as independent across $k = 0, \dots, K$ and distributed standard normal. This is a strong assumption and not necessary for identification and estimation, but is usually assumed in applied work. At the end of this section we discuss some papers that relax this assumption.

The data used to estimate θ will generally have three types of variables. First, quantities of the J products purchased in market t , which are an aggregation of choices made by individual consumers.¹⁸ A market t is implicitly defined by a set of consumers facing the same prices, p_t , characteristics, x_t and demand shocks ξ_t . Aggregate quantities can be converted to market shares by making an assumption about the market size I_t , namely the number of consumer who made choices, including the choice of the outside good. We will then define observed market shares as $s_{jt} = q_{jt}/I_t$. In many applications we think that I_t is sufficiently large so that we can ignore the sampling errors in these shares.¹⁹

Second, we will observe prices, p_{jt} , and ("observed") product characteristics, x_{jt} , of the J products in market t .²⁰ We do not observe all the characteristics consumers observe, but we assume that those that we do observe vary by market and/or product but are common to all consumers in a market.²¹

Third, the data may contain information on consumer demographics D_{ilt} . In micro data, the actual D_{ilt} will be observable. In other data sets the researcher may have access instead to the distribution of demographics, $F_t(D)$ (or have samples from it). At times the researcher will have data that is more aggregated than the consumer-level, but also less aggregated than the market level. For example, the average age of consumers who purchase product j , or the shares by income.

¹⁸In some of the illustrations below we will assume that we observe individual choices, which is often referred to as "micro data". Furthermore, some of the extensions discussed in Section 6 use micro data, but we do not offer a complete treatment of estimation with micro data.

¹⁹At the end of this section, we will revisit the role of sampling error in market shares.

²⁰In some cases the number of products will vary across markets t . For simplicity of notation we focus on $J_t = J$.

²¹In rare cases, one might have characteristics data that varies by consumer. For example, different consumers in the same market might pay different prices when transaction-level data are available. A big issue with such data is that the price paid by a consumer for a purchased product might be observed but data for products not purchased will rarely be observed and need to be imputed.

4.2 What Variation in the Data Can Identify the Parameters?

In the subsection we provide intuition for what variation in the data allows us to identify the parameters θ .²² Our goal is to use the logic of identification to motivate different estimation strategies.

We start by assuming that we have micro data, on individual choices, from a single market and we shut down part of the model by setting $\Sigma = 0$. We show that in this case there is an intuitive two-step procedure for estimating the remaining parameters. This procedure cannot be exactly replicated with market-level data but it provides us a road map for estimation in the more general case. It also provides direction on the type of variation that can identify Γ .

Next, we reintroduce the the random taste shocks, ν_{it} into the model (i.e., we do not restrict Σ to be zero). We show that in order to identify Σ we need different variation than what we use to identify Γ . To illustrate this point we assume we have market-level data from a single market and show the identifying power of moment conditions interacting ξ with IVs.

These simple cases, with data from a single market, are used for illustrative purposes. We use the insight from each of the cases to develop a set of practices that allow the full variation in the data to be leveraged to inform the estimation of θ . In the next section we will bring together these pieces in a general estimation framework that encompasses many of the procedures used in the literature.

4.2.1 Intuition from Individual-level Data

To gain intuition for the more general estimation procedure we start with a simple version of the model and assume that we have individual-level data from a single model. We have two goals in this discussion. First, we propose a simple way to estimate the parameters that will serve as a road map for the more general estimation problem. Second, the discussion will allow us to examine the sources of variation needed to estimate the different parameters.

Assume that we have data $\{y_{ij}, D_i\}_{i=1, \dots, I}$, where $y_{ij} = 1$, for $j = 0, 1, \dots, J$ if consumer i chooses product j and $\sum_{j=0}^J y_{ij} = 1$. All the consumers are from a single market, in the sense that they face the same prices and product characteristics, both observed, \mathbf{x} and unobserved ξ . It might seem impossible to estimate demand in this setting. We only

²²For a more formal treatment of identification see Berry and Haile (2021) in this Handbook.

observe a single snapshot of the market. How could we ever recover how quantities vary with changes in prices if prices do not vary? The answer relies on exploiting variation across households and across products to estimate the choice model, and then using the choice model to compute substitution as we saw in the previous section.

For exposition purposes we shut off certain parts of the model defined by (3.3). Specifically, we will assume that $\Sigma = 0$, namely that heterogeneity will only be driven by observed demographics. Also, we will assume, for this subsection, the price p_j is one of the observed characteristics x_j simply to ease the exposition. Given these assumptions the conditional indirect utility from product j (dropping the subscript t since we have a single market) is given by

$$u_{ij} = \underbrace{x_j \beta_0 + \xi_j}_{\delta_j} + \sum_{k,l} \beta_d^{(l,k)} D_{il} x_{jk} + \varepsilon_{ij}. \quad (4.1)$$

Note, that if we did not have ξ_j then we could estimate the parameters of the model, (β^0, Γ) , by maximizing the likelihood of observing the choices in the sample as a function of x and D_i . The presence of ξ means that we need to modify the estimation somewhat. Specifically, we can estimate the parameters of the model in two steps. In the first step we include a product-specific intercept, that will capture δ , and absorb both $x_j \beta_0$ and ξ_j . In this step we estimate $\tilde{\theta} = (\delta_1, \dots, \delta_J, \Gamma)$ using maximum likelihood. This allows us to “control” for the presence of ξ .

In the second step, we estimate β_0 by “projecting” the estimated $\hat{\delta}$'s on the x 's. If we assume that $E(\xi_j | x_j) = 0$ we can use (weighted) least squares for this second stage. Alternatively, if we are concerned that a subset of the x 's is correlated with ξ we can base the second stage on

$$E(\xi_j | Z_j) = 0, \quad (4.2)$$

where Z are a vector of exogenous variables, which we will discuss further below.

Let us examine the two-step procedure sketched above in further depth. If we parameterize the model according to $\tilde{\theta} = (\delta_1, \dots, \delta_J, \Gamma)$, then it is fairly straightforward to show the first-order conditions implies the maximum likelihood estimates of the product-specific intercepts δ_j are found by setting the observed market shares, or average choice probabilities, equal to the ones predicted by the model. Namely, $\hat{s}_j = \hat{\sigma}(\hat{\delta}_1, \dots, \hat{\delta}_J)$ for a fixed value of Γ . Under quite general conditions (Berry et al., 2013) this relation can be

inverted to yield

$$\hat{\delta}_j = \hat{\sigma}_j^{-1}(\hat{s}_1, \dots, \hat{s}_J). \quad (4.3)$$

As $I \rightarrow \infty$ the limit of this expression will be

$$\delta_j = \sigma_j^{-1}(s_1, \dots, s_J) \quad j = 1, \dots, J,$$

which will play a key role in the estimation with aggregate data, which we discuss below. It can be seen as the limit of (4.3), which comes from the first-order conditions of MLE.

Turning to the first-order conditions with respect to Γ , one can show that maximum likelihood estimates of Γ are the ones that equate the observed and predicted covariance between the demographic variables of those consumers that chose product j and the characteristics of the product. In the limit, Γ is identified as the solution to the system of the $L(K + 1)$ equations

$$E_{Population} [x^k D^l] = E_{Model} [x^k D^l; \Gamma]. \quad (4.4)$$

That is, Γ sets the model's prediction about the covariance between each demographic variable and the product characteristic of the chosen alternative equal to the population counterpart. The MLE moment conditions are simply sample analogues to these limiting moment conditions. As we will discuss below, the intuition gained from (4.4) can be useful even if the researcher does not have consumer choice data.

Suppose we go back to the more general model and allow for unobserved heterogeneity in tastes for characteristics at the individual level. That is, we are back to equation (3.3) where we have both sets of parameters Γ and Σ , and data from a single market. To estimate this model we could consider the same two-step approach as above. The first-order/moment conditions for (δ, Γ) derived above continue to hold. We also have first-order conditions with respect to Σ , which look almost identical to the conditions with respect to Γ . The difference is that it is not clear what is the counterpart of the data covariance that the model is matching, since ν is unobserved. To put it slightly differently, it is not clear what variation or moments in the data identifies these parameters.²³

The answer for how to identify Σ when we observe data from a single market lies in using the variation in the second stage moments defined in equation (4.2). In the next subsection, we show how this moment helps identify Σ .

²³If we observe multiple markets, we can use variation in the choice sets, if it exists, to identify Σ .

4.2.2 The Informational Content of $E[\xi | \mathbf{Z}] = 0$

In this subsection we show how the moment condition given in equation (4.2) provides identifying power for Σ . For exposition purposes, we simplify the model further and assume here that we do not have any demographic variables, D_i , and therefore the choice data is only used to compute aggregate market shares. We therefore can simply assume that we observe market-level data. We assume that we have data from one market and that the indirect utility is given by

$$u_{ij} = \delta_j + \sum_k \beta_\nu^{(k)} \nu_{ik} x_{jk} + \varepsilon_{ij} \quad (4.5)$$

The question becomes what variation in the data pins down the parameter vector (δ, Σ) , where as previously defined Σ is a diagonal matrix with a diagonal equal to $\beta_\nu^{(1)}, \dots, \beta_\nu^{(K)}$

As in the previous subsection, in order to get an expression of the aggregate choice probability that is a function of data and parameters, we can parameterize the δ 's as product-specific intercepts and estimate them as parameters. As before, the estimated $\hat{\delta}$'s will set the predicted shares equal to the observed shares, for any given Σ . In other words, for every Σ there exists $\delta(\Sigma)$ that perfectly explains the observed market shares \hat{s} . There is no information remaining in the (aggregate) choice data alone that would distinguish one set of implied mean utilities $\delta(\tilde{\Sigma})$ from another assignment $\delta(\hat{\Sigma})$. The identification of the true δ requires using more of the structure of the model and adding an additional assumption. We recall that $\delta_j = x_j \beta_0 + \xi_j$.

Berry et al. (1995) propose adding the moment restriction, $E[\xi_j | \mathbf{Z}] = 0$. Below we will discuss variables that might satisfy this condition. This will require that estimates of β_0 and Σ not just fit the aggregate market shares as given by the MLE moments, but also fit the sample analog of this moment condition.

To gain intuition for how this works, let us work with a common exogeneity restriction that $\mathbf{Z} = \mathbf{x} = (x_1, \dots, x_J)$, i.e., \mathbf{Z} is a stack of all the product characteristics in the market.²⁴ Our goal here is not to justify the assumption but rather to understand its empirical usefulness. We can view \mathbf{x} as representing the market structure, namely, a configuration of the number of products and their product positions. The moment restriction (4.2) thus states that the unobserved component ξ_j of mean utility is mean-independent of market

²⁴Observe that this same restriction on product characteristics being determined outside of the model was used in our discussion of Bresnahan (1987) in Section 2.

structure. In particular, the empirical bite of the assumption is that the ξ_j must be uncorrelated with the proximity of competition.

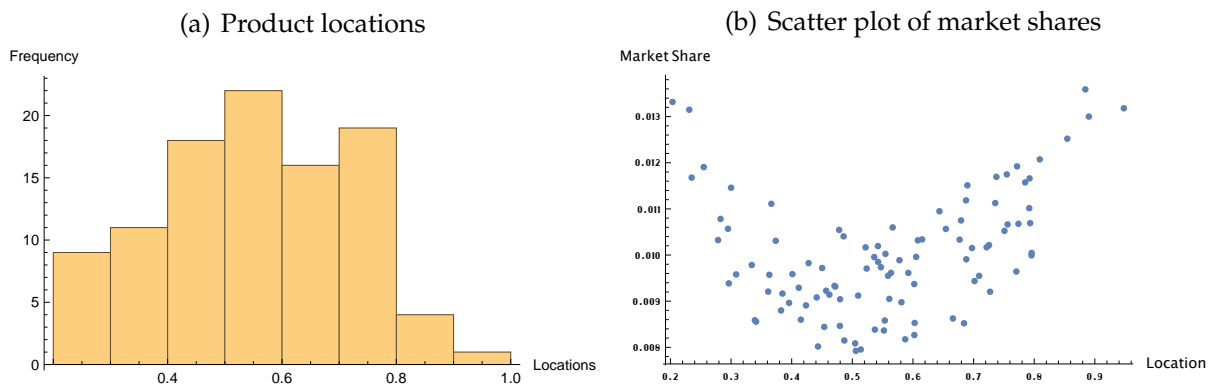
To see this point better, consider a one-dimensional (Hotelling-like) variation of our demand model. The utility to consumer i for product j in this model is

$$u_{ij} = \theta \cdot d(t_i, x_j) + \xi_j + \varepsilon_{ij} \quad j = 0, \dots, J$$

where θ is the travel cost and d is the distance between the location $t_i \in [0, 1]$ of consumer i and the location $x_j \in [0, 1]$ of product j . ξ_j is the mean quality of product j in the population of consumers and ε_{ij} are i.i.d. idiosyncratic taste shocks around this mean quality drawn from a type-1 extreme value distribution.

If $\theta > 0$, i.e., travel costs are positive, a product j will draw demand in higher proportion from other products k that have characteristic x_k close to x_j . The larger the travel cost, the more this “local competition” effect will dominate the substitution patterns from the simple Logit component $\xi_j + \varepsilon_{ij}$ of the model. Suppose we observe data $\{s_j, x_j\}_{j=1}^J$ and we want to infer the magnitude of travel costs θ from the data. To visualize the empirical problem, we plot in Figure 4.1 data generated by drawing 100 products where each product has a location $x_j \in [0, 1]$ and quality $\xi_j \in \mathbb{R}$.²⁵

Figure 4.1: Distribution of product locations and market shares



Notes: The data for these simulations was generated by drawing 100 products from the Hotelling-like model described in the text, where each product was drawn independently from a Beta distribution with both shape parameters equal to 4. In panel (a) we display the histogram of product locations on the line and in panel (b) we show the scatter plot of market shares and product locations in the resulting data.

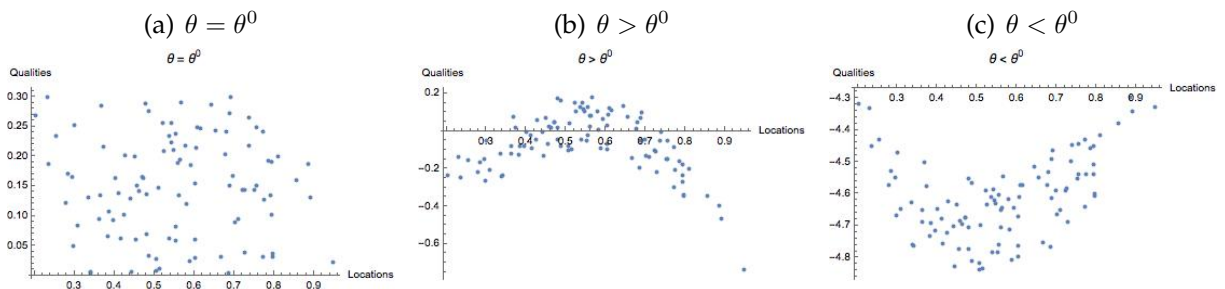
²⁵Each product location was drawn independently from a Beta distribution with both shape parameters equal to 4.

In panel (a) we display the histogram of product locations on the line. The histogram shows that there is relatively more “bunching” of products in this market near the center of the line, and relative isolation of products near end points. We then compute market shares for each product $j = 1, \dots, 100$ based on a true $\theta^0 = 2$ where $\theta^0 \in \Theta$ denotes the true value of travel costs. In panel (b) we show the scatter plot of market shares and product locations in the resulting data. We can see a distinct pattern arise - in the more crowded part of product space (near the middle of the line in this simulation) market shares tend to be relatively smaller. How can we explain this pattern? Why do market shares fall in the middle of the line where products locations are also more tightly concentrated? Intuitively, there are at least two conflicting hypotheses that can explain the correlation of markets shares and locations. One hypothesis is that travel costs are relatively large, and therefore products mostly compete locally. In this case, products that are located in a more crowded part of the line have lower market shares because these products face more competition for the same consumers.

An alternative hypothesis is that travel costs are zero ($\theta = 0$), and products j that are located in a more crowded center part of the line have lower market shares because these products have systematically lower qualities ξ_j .

Market share data alone cannot sort out these alternative explanations, because for any value of θ the implied quality will adjust to match the observed market share. However, assuming (4.2) will do the job. We visually show this in Figure 4.2. The figure shows a scatter plot of locations x_j and implied quality $\xi_j(\theta)$ under three scenarios for travel costs: in panel (a) $\theta = \theta^0$, in panel (b) $\theta > \theta^0$, and in panel (c) $\theta < \theta^0$.²⁶

Figure 4.2: Scatter plot of product locations and quality assignment



Notes: The figure shows a scatter plot of locations x_j and implied quality $\xi_j(\theta)$ under three scenarios for travel costs: in panel (a) $\theta = \theta^0 = 2$, in panel (b) $\theta = 4$, and in panel (c) $\theta = 0$.

²⁶We use $\theta = 0$ (the Logit case) and $\theta = 4$ (a doubling of the true value $\theta^0 = 2$) as our two departures from the true value.

Panel (a) of Figure 4.2 shows that at the true parameter value there is no correlation between quality and location. Panels (b) and (c) show that when θ is different from the truth the data exhibit correlation. In panel (b) the travel cost is overstated and therefore local competition is overstated. Hence to explain the observed market shares the implied quality $\xi_j(\theta)$ has to be systematically higher in the crowded part of product space. In panel (c) the pattern is reversed.

Thus, by using the sample analog of the moment restriction (4.2) in estimation we are shutting off the opportunity for the model to explain the data through a systematic correlation between ξ_j and the local market structure across products j . As we will see below, local market structure measures serve as IVs to estimate the non-linear parameters of the model. In addition to helping us pin down Σ and the implied $\delta(\Sigma)$, the moment conditions implied by (4.2) will help us pin down β_0 .

The key lesson from the above example is that as long we have an assumption like (4.2), the parameters of the model are in principle identified from a cross section of products within a market. In reality, we might need variation across markets to have more power. The key empirical challenge, which we will discuss below is how to choose the IVs that are informative. We discuss this as well as computation and other empirical details for the more general case in the next section.

4.3 The General Estimation Procedure

We now generalize the setting described in the previous section to allow for multiple markets $t = 1, \dots, T$. We continue to use a conditional moment restriction $E[\xi_{jt} | \mathbf{Z}_t] = 0$, where \mathbf{Z}_t is a vector of exogenous variables, as the basis for estimation of heterogeneous preferences in the general Mixed Logit demand model. In Section 4.2.2 we saw the identifying power of this moment condition, when we focused only on within-market variation across products. We now discuss how to use this moment condition in practice in the general setting that has both within- and across-market variation. As we discuss below, having multiple markets potentially provides additional variation in choices set and characteristics of products, and therefore helpful in estimation.

We note that the conditional moment implies a large set of potential IVs $z_{jt} = A_j(\mathbf{Z}_t) \in \mathbb{R}$ for which the unconditional moment restriction holds

$$E[z_{jt}\xi_{jt}] = 0. \tag{4.6}$$

We refer to A_j as the IV function, and z_{jt} as IVs. We provide some guidance on the choice of IVs below. For a given choice of IVs, estimation proceeds on the basis of empirical analogues of the population moments (4.6) for each IV n . Let

$$m^n(\theta) = \hat{E}[z_{jt}^n \xi_{jt}(\theta)] \quad (4.7)$$

be the n^{th} moment (where $\hat{E}[\cdot]$ is the expectation taken with respect to the sample distribution), and θ are the parameters of the model. Stacking the moments as $\mathbf{m}(\theta) = (m^n(\theta))_{n=1}^N$, the standard GMM estimator for the problem as formulated by Berry (1994) and Berry et al. (1995) is

$$\hat{\theta} = \arg \min_{\theta} \mathbf{m}(\theta)' W \mathbf{m}(\theta) \quad (4.8)$$

for a positive definite $N \times N$ weighting matrix W . Inference can be based on the standard tools for GMM.²⁷

A key practical difficulty in using this moment condition is computing ξ_{jt} , and therefore the moment, as a function of data and parameters. This is where the logic discussed in Section 4.2.1 enters. Specifically we will solve a non-linear system of equations like (4.3) for each market t in the data - first the mean utility vector δ_t must be inverted from market shares in each market, and the econometric error $\xi_{jt}(\theta)$ then computed. Berry (1994) and Berry et al. (1995) propose a contraction mapping algorithm that globally converges and can be used for computation. In Section 4.3.4 we discuss this and other algorithms in greater detail.

4.3.1 Instrumental Variables

The above discussion assumed that we have exogenous variables \mathbf{Z}_t that can be used as the basis for constructing IVs such that (4.2) holds. We now discuss various variables that have been used in the literature for these purposes. Before we do so we review the dual role of IVs. This is best done with a series of examples. We start with the Logit model, which was described in Section 3.2. For the Logit model $\sigma_j^{-1}(\mathbf{s}_t, \mathbf{x}_t, \mathbf{p}_t) = \ln(s_{jt}/s_{0t})$, and the estimating equation is

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \delta_{jt} \equiv x_{jt}\beta_0 + \alpha_0 p_{jt} + \xi_{jt}. \quad (4.9)$$

²⁷See Freyberger (2015) and Berry et al. (2004) for more detail.

This equation can be estimated using linear methods. In this case, an IV is needed if one of the right-hand side variable, say p_{jt} , is correlated with the error term. The intuition for what constitutes a good IV is just the “standard” logic in linear models. We will provide examples below.

This particular role of the IV remains even when estimate models that allow for more heterogeneity in preferences. However, now there is an additional role that links directly to our discussion in Section 4.2.2. To see this consider the Nested Logit model. As Berry (1994) shows, the inversion is now given by $\sigma_j^{-1}(\mathbf{s}_t, \mathbf{x}_t, \mathbf{p}_t) = \ln(s_{jt}/s_{0t}) - \rho \ln(s_{jt}/s_{G(j)t})$, and the estimating equation is

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = x_{jt}\beta_0 + \alpha_0 p_{jt} + \xi_{jt} + \rho \ln\left(\frac{s_{jt}}{s_{G(j)t}}\right) \quad (4.10)$$

where $s_{G(j)t}$ is the share of nest G , and ρ is the nesting parameter. This model, like the Logit model, can also be estimated using linear methods. Relative to equation (4.9), which describes the estimation equation of the Logit model, in the Nested Logit model we have an additional term: the within nest share. This last term is a function of the share of product j , s_{jt} , and therefore will be correlated with ξ_{jt} . In other words, even if we believe that $E(\xi_{jt}|x_{jt}, p_{jt}) = 0$ we still cannot consistently estimate this equation using OLS: we need an IV.²⁸

This formulation also suggests where we could find IVs. We want variables that generate variation in the nest share but that also satisfy (4.2). Natural candidates are variables that impact the other products in the nest, and therefore the nest share, but are uncorrelated with ξ_{jt} . For example, if looking at a given market over time we might want to use the entry and exit of products into the nest.

As we move to models that allow for even richer patterns of heterogeneity we generally do not have an analytic expression for the inversion. However, the intuition from above continues to hold. Gandhi and Houde (2020) and Gandhi et al. (2021) show that $\sigma_j^{-1}(\mathbf{s}_t, \mathbf{x}_t, \mathbf{p}_t) = \ln(s_{jt}/s_{0t}) - f_j(\mathbf{s}_t, \mathbf{x}_t, \mathbf{p}_t)$, where $f_j(\cdot)$ is an unknown function that in the BLP algorithm (discussed in Section 4.3.4) is computed numerically. In Section 4.4.2 we discuss ways to approximate $f_j(\cdot)$.

We now turn to a discussion of what exogeneity restrictions and IVs researchers have used in practice.

²⁸As we noted on page 21, the Nested Logit model can be viewed as a special case of the Mixed Logit model. In the above formulation of the model ρ is the equivalent of Σ in our general formulation.

BLP Instruments By far, the most popular IVs are $Z_t = \mathbf{x}_t$, namely the characteristics of all products in the market.²⁹ Typically price, and maybe advertising, will be excluded from this set. The identifying power of this exogeneity restriction is based on the same logic that we saw in Sections 2 and 4.2.2. They are informative because they can be used to measure the proximity of competition, just as we saw in Figure 2.1, and therefore should be correlated with price and other endogenous variables. They will also be correlated with terms like the within nest share, in the Nested Logit model, or $f_j(\mathbf{s}_t, \mathbf{x}_t, \mathbf{p}_t)$ in the more general model.

The question still remains which of the many possible functions $A_j(\cdot)$ of \mathbf{x}_t should we use to construct IVs. Different suggestions have been made in the literature. Berry et al. (1995) propose to use: (1) own product characteristics, \mathbf{x}_{jt} , (2) sum of characteristics of the other products produced by the same firm (for multi-product firms), $\sum_{j' \neq j, j' \in \mathcal{J}_{f(j)}} \mathbf{x}_{j't}$, and (3) sum of the characteristics of competitor products $\sum_{j' \notin \mathcal{J}_{f(j)}} \mathbf{x}_{j't}$. The logic for this particular set is as follows. The own product characteristics are instruments "for themselves." The other two sets try to capture the logic that the price of product j (and $f_j(\cdot)$) will depend on characteristics of other products, and that the dependence differs if these are own products or competitors products.³⁰

Gandhi and Houde (2020) propose to refine how we use the information in (4.6) in order to improve empirical performance and avoid weak IV challenges that can arise in practice. Specifically, they use the economic structure of the model to motivate a class of IVs they term "differentiation Instruments", which intuitively capture the relative isolation of each product in characteristics space. Defining $\mathbf{d}_{jkt} = \mathbf{x}_{jt} - \mathbf{x}_{kt}$ as the vector of characteristic differences between product j and product k in market t , they construct two distinct sets of differentiation IVs that are useful for applied work.

The first set consists of (1) own product characteristics, \mathbf{x}_{jt} , (2) the distance squared between product j and other products along dimension k , $\sum_{j' \neq j} (d_{jj't}^k)^2, \forall k$ and (3) the interaction between the distance in dimensions k and l , $\sum_{j' \neq j} d_{jj't}^k \times d_{jj't}^l, \forall k \neq l$. The sum of square of characteristic differences captures a continuous measure of product isolation proportional to the Euclidean distance of product j along each dimension k . The interaction terms capture the covariance between two dimensions of differentiation.

²⁹One of the reasons for the popularity of these IVs is that they typically do not require any additional data: they are part of the data needed to estimate the model to start with.

³⁰There is some disagreement among researchers about whether the term "BLP Instruments" refers narrowly to these specific functional forms or more broadly to the idea $E[\xi_{jt} | \mathbf{x}_t] = 0$.

A second set consists of: (1) own product characteristics, \mathbf{x}_{jt} , (2) the number of products within a certain “band” of j , $\sum_{j' \neq j} 1(|d_{jj't}^k| < \kappa^k)$, $\forall k$, and (3) the interactions between the number in dimensions k and l , $\sum_{j' \neq j} 1(|d_{jj't}^k| < \kappa^k) \times 1(|d_{jj't}^l| < \kappa^l)$, $\forall k \neq l$. These IVs try to capture the economics behind models of localized competition. The second element measures the number of “close-by” products along each dimension of differentiation. The interaction of the indicator function with $\mathbf{d}_{jj't}$ captures the correlation in characteristics between firms that are direct competitors. When characteristics are discrete, the indicator variables can be replaced by $1(d_{jj't}^k = 0)$; which can be thought of as a product-segment indicator. Moreover, additional neighborhoods can be constructed to impose additional restrictions on the model (e.g. $0 < |d_{jj't}| \leq \kappa_1$, $\kappa_1 < |d_{jj't}| \leq \kappa_2$, etc.)

All these various permutations of IVs are motivated by the search for “powerful” IVs, assuming the IVs are valid, namely that $E(\xi_{jt} | \mathbf{x}_t) = 0$. It is not difficult to come up with economic models where this validity is violated. For example, if characteristics are chosen by the firms after they observe (some components of) ξ_t then this assumption will be violated. A typical defense of the assumption is that even if the characteristics are chosen, they are chosen in advance and before ξ_t is observed. For example, in the case of cars elements of design are chosen many years in advance.

This is only a partial defense since firms might be forward-looking and could anticipate in part the realization of ξ_t . This can be dealt with, if we observe panel data, by relying on the ideas of the dynamic panel literature (Arellano and Bond, 1991; Blundell and Bond, 1998). For example, Sweeting (2013) assumes that $\xi_{jt} = \rho \xi_{jt-1} + u_{jt}$ where u_{jt} is unanticipated at time $t - 1$. He then bases the estimation on the conditional moment

$$E[\xi_{jt} - \rho \xi_{jt-1} | \mathbf{x}_{t-1}] = 0. \quad (4.11)$$

Hausman Instruments The “textbook” IVs for prices when estimating demand are cost variables. In most IO applications cost is not observed. Furthermore, even if we observe (marginal) cost, or some proxies for it, rarely will it vary by product.³¹ Hausman et al. (1994) and Hausman (1996) propose using prices in other markets as IVs, often called “Hausman IVs”. Nevo (2001, 2000a) builds on this idea to estimate a discrete choice model. To see how these IVs operate, consider estimation of equation (4.9) and assume that p_{jt} is correlated with ξ_{jt} . The idea is to use prices in other markets, namely, $p_{jt'}$, $t' \neq t$,

³¹Villas-Boas (2007) uses cost IVs by gathering information on input prices and interacting these prices with product dummy variables. This is trying to capture the idea different products use a different mix of inputs and therefore will have a different relationship between prices and input prices.

as IVs. Depending on the structure of the data we could use all $t' \neq t$, or markets in the same time period, same region, or otherwise matched. These IVs are potentially valid if, conditional on x_t and $x_{t'}$, pricing is independent across markets and ξ_{jt} and $\xi_{jt'}$ are independent. These IVs are trying to exploit common cost shocks across markets for identification.

There are two main problems with these IVs. First, it is not difficult to come up with arguments why they are not valid. For example, if there is an (unobserved) promotional or advertising campaign across markets then the independence assumption would be violated. Second, it is less obvious how the prices of the own brand in other markets will help, for example, in estimation of equation (4.10): it is not clear that these IVs will be correlated with the within nest share.³² In principle, one could use $p_{j't'}$, $j' \neq j$, and $t' \neq t$, i.e., the price of other products in other markets, to proxy for cost shocks to other products. However, we are unaware of a published paper that uses this approach.

Waldfoegel Instruments In some cases researchers have used attributes of other markets, such as demographics, in a slightly different way than Hausman instruments. For example, Town and Liu (2003) estimate the welfare associated with the Medicare HMO program known as Medicare+Choice. To do so they estimate a Nested Logit model at the county-level and use the fact that each plan is typically offered in several counties. One of the IVs they use is the mean number of competitors in the other counties where the plan is offered. Similarly, Fan (2013), when estimating demand for newspapers using county-level data, exploits the fact that newspapers sell in multiple counties and uses demographics in other counties as IVs.

Like Hausman IVs, these IVs use information in other markets, but the logic is different: Hausman IVs rely on common cost shocks, while these IVs rely on consumption or preference externalities. If the product is offered in multiple counties the price and characteristics of the product will be impacted by the attributes, say demographics, in the other counties. So, for example, if a product is offered in counties A and B its price should be a function of demographics in both counties. For this reason these IVs are often referred to as "Waldfoegel IVs" (Waldfoegel, 2003). These IVs are valid if, conditional on the variables included in the model, ξ_{jt} is not correlated across counties, just like the requirement for the Hausman IVs. For the same reason this assumption was suspect there,

³²As Berry and Haile (2014) show, separate IVs are needed for prices relative to market shares - the model has two distinct sets of endogenous variables (as we saw in the Nested Logit example).

it could be suspect here as well. Furthermore, the set of counties covered by a plan is not exogenous and could be an indication that the counties are similar in some ways.

4.3.2 Additional Sources of Variation

We now briefly discuss additional sources of variation that can aid in estimation and identification.

Multiple markets Some of the above IVs discussed above could in principle be constructed with data from a single market, e.g., BLP IVs (as we saw in Section 4.2.2. However data from multiple markets can significantly aid in identification and estimation.

In the case of BLP IVs a main advantage of having multiple markets is the potential for variation in the number of products and their characteristics. For example, consider the estimation of equation (4.10) using data across markets. The within nest share might vary because different products are available in different markets. This has been found to be a powerful way to estimate this model, especially in cases where entry and exit of products are arguably exogenous. This idea generalizes. The intuition given in Section 4.2.2 that focused on the informational content of equation (4.2) with data from a single market, can be extended to multiple markets, especially as competitive conditions vary across markets.

An additional way data from multiple markets can be used is through demographic data. As we saw in Section 4.2.1 having consumer level data can aid in identification and estimation. Data from multiple markets, with variation in demographics can achieve similar results. For example, suppose that we observe markets with different distribution of ages. This allows us to correlate the outcomes and demographics. We can do that by imposing "micro moments" as we discuss below. Or by using the full distribution of demographics to compute the shares in equation (3.7). See Nevo (2000b) and the computational discussion below for details.

Furthermore, having data from multiple markets allows us to control for unobserved product characteristics that do not vary across markets. Finally, Hausman IVs and Wald-fogel IVs require multiple markets in order to be computed. In sum, data from several markets is very helpful in estimation of the model, and in general the more the better.

Micro moments and Second Choice Data As we saw in Section 4.2.1 having information on demographics and consumer choices can be very useful in estimation. The in-

tuition gained from equation (4.4) can be useful even if the researcher does not have consumer choice data (i.e., data that includes both the choices of individuals and their demographics). These moments can be computed from other data sources and added to the estimation. For example, suppose a researcher is estimating demand for cars and has information on the average family size conditional on owning a minivan. We can match the model’s prediction of this choice behavior in the population with the sample analogues. This mimics the logic of the moments in equation (4.4), “as if” we had micro data.

Petrin (2002) follows this approach and finds that the micro moments impact the estimates of consumer heterogeneity and have important implications for his estimate of the welfare gains from the introduction of minivans. Follow-up research has found similar results. It is therefore advisable to add micro moments whenever possible.³³

Another source of data that is powerful, albeit much harder to obtain, is second choice data. Berry et al. (2004) use survey data on second choices, i.e., what the consumer would choose if the actual chosen alternative were not in the choice set. Such data provide a direct empirical insight into substitution patterns among products and a useful source of identification and estimation.

Supply-Side Moments Another way to help identify the parameters is to add supply-side moments. Assume that the marginal cost is given by

$$mc_{jt} = \mathbf{w}_{jt}\boldsymbol{\gamma} + \omega_{jt},$$

where \mathbf{w}_{jt} is a vector of observed characteristics of product j , ω_{jt} is an unobserved component, and $\boldsymbol{\gamma}$ is a vector of parameters to be estimated. If we further assume the Nash-Bertrand pricing model discussed in Section 2.1.1 and combine the cost with the pricing equation (2.2) we get

$$\mathbf{p}_t = \mathbf{w}_t\boldsymbol{\gamma} + \Omega^{-1}\mathbf{q}(\mathbf{p}_t) + \boldsymbol{\omega}_t. \quad (4.12)$$

Using this equation we can form supply-side moments by assuming $E(\omega_{jt}|\mathbf{Z}_t) = 0$, where \mathbf{Z}_t is a vector of IVs that includes products characteristics and cost shifters ($\mathbf{z}_{jt} = [\mathbf{x}_t, \mathbf{w}_{jt}]$ and $\mathbf{Z}_t = [z_{1t}, z_{2t}, \dots, z_{Jt}]$.) Note, that this equation is informative about both the supply parameters, $\boldsymbol{\gamma}$ and the demand parameters, which impact Ω . It also suggests that the cost

³³See Grieco et al. (2021) for an efficient estimator when micro data is available.

shifters could be used as additional demand-side IVs. We can combine the demand and supply-side moments and estimate the parameters using GMM.

4.3.3 Efficiency

The conditional moment $E[\xi_{jt} | \mathbf{Z}_t] = 0$ implies a large set of potential unconditional moments $E[z_{jt}\xi_{jt}] = 0$, where $z_{jt} = A_j(\mathbf{Z}_t)$ are IVs, which are the basis for estimation. The choice of IVs is closely tied to the efficiency of the GMM estimator. We can use the semi-parametric efficiency bound (Chamberlain, 1987) to guide strategies for constructing IVs and efficient estimation. There are two basic approaches.

1. We can allow the dimension of the IVs $A_j(\mathbf{Z}_t)$ to grow with the sample size and capture the informational content in $E[\xi_{jt} | \mathbf{Z}_t] = 0$. The asymptotic efficiency bound is reached by applying the optimal weight matrix W^* to the GMM problem with a suitably rich set of “low-order” basis functions as IVs (see, for example, Donald et al. (2003)). The Differentiation IVs above can serve as such a basis class.
2. We can compute the “optimal IVs” from Chamberlain (1987). This approach is especially likely to be productive, if IVs $A_j(\mathbf{Z}_t)$ do not approximate well the full informational content in $E[\xi_{jt} | \mathbf{Z}_t] = 0$. Theoretically, the optimal IVs are given by

$$z_{njt}^* = E \left[\frac{\partial \xi_j(\theta^0)}{\partial \theta} \middle| \mathbf{Z}_t \right] \quad n = 1, \dots, \dim(\theta) \quad (4.13)$$

where θ^0 is the true value of the parameters. This creates a just-identified problem, e.g., as many IVs as parameters and the weighting matrix is the identity matrix. These are obviously not feasible but can be heuristically approximated. For example, Reynaert and Verboven (2014) explore the heuristic

$$E \left[\frac{\partial \xi_j(\theta^0)}{\partial \theta} \middle| \mathbf{Z}_t \right] \approx \frac{\partial \xi_{jt}(s_t, \hat{\mathbf{p}}_t, \mathbf{x}_t; \hat{\theta})}{\partial \hat{\theta}} \bigg|_{\xi_{jt}=0, \forall j,t} = A_j(\mathbf{x}_t). \quad (4.14)$$

Since the IV vector depends on θ , users must first obtain an estimate of the parameters, denoted by $\hat{\theta}$.

The two approaches are ultimately complementary as discussed in Gandhi and Houde (2020) and Conlon and Gortmaker (2020). The performance of the optimal IV approxima-

tion depends critically on good first-stage estimates. In particular if weak IVs are used in the first stage, the approximation will not work well in practice.

4.3.4 Computational Algorithms

Several computational algorithms have been suggested in the literature to solve the estimation problem we discussed above. We focus on three of them.³⁴

Nested Fixed Point Berry et al. (1995) provide a method to compute the estimator they propose. We now describe the basic steps, for more details see Nevo (2000b) and for updated best practices (and code) see Conlon and Gortmaker (2020). The method consists of the following steps.

In a preliminary step, we draw R random draws from $F_\nu(\nu)$, which is the (standardized) parametric distribution assumed for ν (almost always a standard normal), and \hat{F}_D , which is either an estimated parametric distribution (e.g., log normal for income estimated outside the model), or an empirical distribution (for example, from Census data). These draws are held constant throughout the computation. Denote them as $\hat{F} = \left\{ \hat{\nu}_{it}, \hat{D}_{it} \right\}_{i=1}^R$. In most data sets, we will observe quantities and not market shares. Quantities are converted to market shares by making an assumption on the total market size, namely all the consumers who purchase and those who decided not to purchase. Let I_t denote this quantity, then $s_{jt} = q_{jt}/I_t$. With these preliminaries in hand the algorithm proceeds as follows.

1. Step 1: for a given value of the “non-linear” parameters, Γ and Σ , and vector of mean utilities δ_t compute market shares predicted by the model. The easiest way to do this is via simulation using the draws from the preliminary step,³⁵ namely:

$$\tilde{\sigma}(\delta_t; \Gamma, \Sigma, \mathbf{x}_t, \mathbf{p}_t, \hat{F}) = \frac{1}{R} \sum_{i=1}^R \frac{\exp\{\delta_{jt} + (x_{jt}, p_{jt}) \cdot (\Gamma D_{it} + \Sigma \nu_{it})\}}{1 + \sum_{k=1}^J \exp\{\delta_{kt} + (x_{kt}, p_{kt}) \cdot (\Gamma D_{it} + \Sigma \nu_{it})\}}.$$

2. Step 2: For for a given value of the “non-linear” parameters, Γ and Σ , compute the vector of mean utilities that equates the shares predicted in Step 1 to those observed in the data. This can be computed by starting with a guess for δ_t (say the values

³⁴See Hong et al. (2021) for an additional method that we do not discuss here.

³⁵See Nevo (2000b) and Conlon and Gortmaker (2020) for alternative, more efficient ways to compute these shares.

from the Logit model $\ln(s_{jt}/s_{0t})$ and computing the contraction mapping proposed by Berry (1994):

$$\delta_t^{r+1} = \delta_t^r + \ln(s_t) - \ln \tilde{\sigma}(\delta_t^r; \Gamma, \Sigma, \mathbf{x}_t, \mathbf{p}_t \hat{F}).$$

Stop when $\|\delta_t^r - \delta_t^{r-1}\| < \tau$, where τ is a pre set tolerance level (say 10^{-12}).³⁶

3. Step 3: Use the result of step 2, $\delta_{jt}(\Gamma, \Sigma)$, to compute $\xi_{jt} = \delta_{jt}(\Gamma, \Sigma) - x_{jt}t\beta_0 - \alpha_0 p_{jt}$, interact it with the IVs to form the GMM objective function and compute (4.8) using a non-linear search routine.

This algorithm is relatively easy to program, although to improve computational speed various bells and whistles are needed. See Nevo (2000b) and Conlon and Gortmaker (2020) for details and code.

Mathematical Programming with Equilibrium Constraints (MPEC): Dubé et al. (2012) advocate the use of an MPEC algorithm instead of the above Nested fixed point. The basic idea is to maximize the same GMM objective function as above subject to the constraints that the predicted shares equal the observed shares. However, demand shocks, ξ are treated as parameters. Formally,

$$\begin{aligned} \min_{\theta, \xi} \quad & \xi' z W z' \xi \\ \text{subject to} \quad & \tilde{\sigma}(\delta(\xi); \mathbf{x}, \mathbf{p}, \hat{F}, \theta) = \mathbf{s} \end{aligned}$$

Note that both θ and ξ in this problem and therefore the search is over a much higher dimension search than (4.8). ξ is now a vector of parameters, and unlike before it is not a function of θ . The advantage of this approach is that it avoids the need to perform the inversion at each and every iteration of the search. This inversion can be a significant computational cost, especially when performed for values of the parameters far from θ^0 . The resulting programming problem can be quite large, but there are off-the-shelf programs (e.g., Knitro) that can solve it effectively. Dubé et al. (2012) report significant speed improvements over the nested fixed point. Note, that this is purely a computational algorithm: the result should be identical to the result of the BLP algorithm.

This approach is more complicated to program, and to get the computational benefits one needs to analytically provide various derivatives. Once programmed properly

³⁶With a tolerance level that is not strict enough the algorithm can become unstable as shown by Knittel and Metaxoglou (2014).

it seems to perform well, but some have found it slow in very large problems (many markets and many products) and not worth the extra programming time.

Approximate BLP (ABLP): Lee and Seo (2015) propose an alternative estimator with significant computational advantages that they call Approximate BLP. The basic idea is to approximate the share equation $\sigma(\cdot)$ using a first-order Taylor approximation. This allows them to substitute an analytic inversion, for the numerical inversion, in Step 2 of the original BLP algorithm. Like the MPEC algorithm the inversion is exact only at the solution, but unlike MPEC the optimization is over a lower-dimensional parameter space.

They compute a first-order Taylor approximation to $\tilde{\sigma}(\boldsymbol{\xi}_t; \mathbf{x}_t, \mathbf{p}_t, \hat{F}, \theta) \equiv \tilde{\sigma}(\boldsymbol{\delta}(\boldsymbol{\xi}_t); \mathbf{x}_t, \mathbf{p}_t, \hat{F}, \theta)$ around $\boldsymbol{\xi}_t^0$ given by

$$\ln \tilde{\sigma}(\boldsymbol{\xi}_t; \mathbf{x}_t, \mathbf{p}_t, \hat{F}, \theta) \approx \ln \tilde{\sigma}^A(\boldsymbol{\xi}_t; \mathbf{x}_t, \mathbf{p}, \hat{F}, \theta) \equiv \ln \tilde{\sigma}(\boldsymbol{\xi}_t^0; \mathbf{x}_t, \mathbf{p}_t, \hat{F}, \theta) + \frac{\partial \ln \tilde{\sigma}(\boldsymbol{\xi}_t^0; \mathbf{x}_t, \mathbf{p}_t, \hat{F}, \theta)}{\partial \ln \boldsymbol{\xi}_t'} (\boldsymbol{\xi}_t - \boldsymbol{\xi}_t^0).$$

They equate this approximation to the observed shares, and invert this relation to get

$$\boldsymbol{\xi}_t = \Phi_t(\boldsymbol{\xi}_t^0, \theta) \equiv \boldsymbol{\xi}_t^0 + \left[\frac{\partial \ln \tilde{\sigma}(\boldsymbol{\xi}_t^0; \mathbf{x}_t, \mathbf{p}_t, \hat{F}, \theta)}{\partial \ln \boldsymbol{\xi}_t'} \right]^{-1} (\ln s_t - \ln \tilde{\sigma}(\boldsymbol{\xi}_t^0; \mathbf{x}_t, \mathbf{p}_t, \hat{F}, \theta)).$$

This analytic inversion (of the approximation) allows them to skip the numerical inversion in Step 2 of BLP.

With the aid of the approximation they can estimate the parameters θ by

$$\min_{\theta} \Phi(\boldsymbol{\xi}^0, \theta)' \mathbf{z} \mathbf{W} \mathbf{z}' \Phi(\boldsymbol{\xi}^0, \theta)$$

They nest this idea into the following procedure. Guess $\boldsymbol{\xi}_t^0$ and set $r = 1$

1. Step 1: Compute a GMM estimate

$$\theta^r = \arg \min_{\theta} \Phi(\boldsymbol{\xi}^{r-1}, \theta)' \mathbf{z} \mathbf{W} \mathbf{z}' \Phi(\boldsymbol{\xi}^{r-1}, \theta)$$

2. Step 2: Update $\boldsymbol{\xi}$

$$\boldsymbol{\xi}^r = \Phi(\boldsymbol{\xi}_t^{r-1}, \theta^r)$$

and $r = r + 1$

3. Repeat Steps 1 and 2 until convergence.

Lee and Seo show this estimator is equivalent to the BLP estimator in large samples. The advantage of this approach is that like MPEC it avoids inversion at each stage, but has low-dimensional parameters search.

4.4 Extensions

4.4.1 Error in Market Shares

We typically assume that aggregate quantities are based on a large number of underlying choices and therefore measured without error. This assumption can be problematic when products have a small market share: with a large number of products even with samples generated from thousands of consumers market shares can be measured with error. This is especially problematic if the data include a large number of products with a market share of zero. Ad hoc fixes, sometimes used in practice, such as dropping zeros from the data or replacing them with small positive numbers, are subject to biases which can be quite large. Mathematically, the root of the problem is that the slope of the inverse demand function approaches infinity as the share approaches zero.

Solutions to this problem can be split into two groups, depending on how we view the root for the problem. Which approach is more appropriate for a specific data set depends on how the zeroes are believed to be generated in that data set. The first group views the root of the problem as the wedge between choice probabilities, which come from the theoretical demand model, and market shares, which are the empirical estimates based on the realized choices of consumers in the data. Although the choice probabilities are strictly positive in the underlying model, observed shares may be zero due to sampling error. This is more likely if the underlying choice probability is small.

Gandhi et al. (2021) take this approach. At a high level, they construct lower and upper bounds for the inverse demand function by adding a bit of noise to the observed shares. If they observe a set of products whose empirical shares are unlikely to be zero, then they can point identify the parameters. If there are no such products (for example, because the number of consumers is small), their bound construction leads to a set of moment inequalities that partially identify the parameters. They apply this approach to scanner data and find that the new approach yields demand estimates that can be more than twice as elastic as standard estimates that select out the zeroes.

Dubé et al. (2021) tackle the zeroes problem from a different angle. They assume that $s_{jt} = 0$ if and only if product j is not in the set of products that the consumers in market t consider, or in other words, the choice set. They then offer a specific model of the selection into the choice set in order to estimate the model. Their results rely on carefully placed separability and exclusion restrictions.

4.4.2 Non-parametric and Flexible Estimation

Up to this point we have assumed a parametric functional form for utility, given in (3.3), and specific distribution of heterogeneity. The identification results in Berry and Haile (2014) hold for more general models. Therefore, an obvious question is whether more flexible models, that imply flexible substitution patterns, can be estimated.

In order to appreciate the problem of flexible estimation, it is useful to recall the integral (3.7) that defines aggregate demand in the Mixed Logit model. Up until now in our discussion we have treated the distribution of consumer “types” in the population $F(D_{it}, \nu_{it})$ as a known distribution: we assume we have data to estimate the distribution of D_{it} , and we assumed a parametric distribution for ν_{it} . A more flexible model is to keep the type-1 extreme value distribution assumption on ε_{ijt} , but allow for a flexible mixing distribution. The joint distribution $F(D_{it}, \nu_{it})$ is an unrestricted distribution that is estimated from the data. In this case (3.7) can be treated as an integral equation for identifying F .³⁷ ³⁸ A further generalization maintains the linear utility form in equation (3.3) but treats $(\alpha_i, \beta_i, \varepsilon_{it})$ as distributed according to a general distribution $F(\alpha_i, \beta_i, \varepsilon_{it})$. In this case the integral equation becomes

$$s_{jt} = s_j(\mathbf{x}_t, \mathbf{p}_t, \boldsymbol{\xi}_t) = \int \mathbf{1}(u(x_{jt}, p_{jt}, \xi_{jt}; \alpha_i, \beta_i, \varepsilon_{ijt}) > u(x_{kt}, p_{kt}, \xi_{kt}; \alpha_i, \beta_i, \varepsilon_{ikt}), \forall k \neq j) dF(\alpha_i, \beta_i, \varepsilon_{it}; \theta). \quad (4.15)$$

The question in either (3.7) or (4.15) is whether the distribution F can be estimated in a flexible way, e.g., either non-parametrically or with a flexible parameterization θ . Un-

³⁷The joint distribution $F(D_{it}, \nu_{it})$ can be constrained so that the marginal F_D equals the actual distribution of demographics in the market, which may be observed/known.

³⁸With micro data the problem is different. For example, Dubois et al. (2020) utilize a long panel of consumer choices to estimate a Logit model where the coefficients on price and characteristics are estimated separately for each consumer, which avoids making distributional assumptions on the random coefficients.

fortunately the presence of ξ_{jt} in the integral equation (4.15) complicates the application of standard estimators for flexible heterogeneity in discrete choice models.³⁹

One approach to the problem is to change the focus away from estimation of preferences to estimation of demand by estimating the demand function σ (or more precisely, $\sigma^{-1}(\cdot)$) directly in a flexible way. The basic idea is to approach estimation of the inverse demand function $\sigma^{-1}(s_t, p_t, x_t)$ directly rather than estimating a model of preferences first as a means to constructing demand. These approaches all have to address the dimensionality problem of $\sigma^{-1}(\cdot)$ that arises without an explicit preference structure - the model is now expressed in product space and as we discussed in Section 3.1, the number of parameters to estimate can be very large.

Compiani (2019) is an example of this approach. He proposes to directly non-parametrically estimate the inverse demand function $\sigma^{-1}(\cdot)$ through a sieve approximation. This has the advantage of requiring fewer assumptions than aggregating demand from a random coefficient choice model - in principle it only requires invertibility of demand that is guaranteed by the connected substitutes condition in Berry et al. (2013). Thus, the class of demand models that are consistent with the estimator is broader than Mixed Logit models, including models with some degree of product complementarity as well as models that allow for behavioral economic effects at the level of the consumer. The cost is that there is a curse of dimensionality encountered in a product space model, which was the motivation for using explicit choice models in the literature as discussed above. Compiani (2019) shows that Bernstein polynomials can allow for some parsimony to be added to the problem through linear constraints on the parameters that are motivated by theoretical restrictions on demand implied by choice models, such as monotonicity and symmetry. The specification nevertheless requires significant data for moderately sized markets.

Another example is Salanié and Wolak (2019) who use the idea of “artificial regressions” from Davidson and MacKinnon (1989) that takes a first-order expansion of the residual function used in non-linear IV estimation, e.g., a first-order expansion of $\xi_{jt}(\theta)$ around an initial value for $\theta = \theta'$. The approach retains the full parsimony of the underlying Mixed Logit model, and is “fast” in the sense of being a single IV regression, however it is approximate in that it only iterates the regression a single time (multiple iterations would equate to a Gauss-Newton optimization of the GMM objective function) and also

³⁹See e.g., Fox and Gandhi (2016), Fox et al. (2011), and Fox et al. (2016) for a discussion of identification and estimation of discrete choice models with flexible heterogeneity where product-market unobservables ξ_{jt} are not present.

requires a starting value to run. It is a fast way however to generate starting values that can be used for non-linear estimation.⁴⁰

A third example is Fosgerau et al. (2020) who generalize the analytic structure of the Nested Logit functional form for inverse demand to a broader class of consumer demand models - what they call “inverse product differentiation logit” model. The approach is also based on linear IV estimation, and the generalized term they add to the model allows for complementarity among products as well as being consistent with an underlying representative agent model of consumer demand. However it is a preference model expressed in “product space” as opposed to a characteristics space model, and requires the researcher to specify ex-ante dimensions of segmentation in the market where products can be categorized.

A challenge with the above approaches is that while they estimate the demand function $\sigma(\cdot)$ in a flexible way under different conditions, they do not estimate the distribution F of consumer heterogeneity that is a central to many applications. Gandhi et al. (2021) take a different approach that allows researchers to flexibly recover both demand $\sigma^{-1}(\cdot)$ and the distribution of heterogeneity F . They proceed in two steps. In the first step, like the above papers, they estimate the inverse demand model. Their main specification, does not directly tie the inverse demand to preference model parameters, unlike Salanié and Wolak (2019) and Fosgerau et al. (2020). However, unlike Compiani (2019) they use the structure of Mixed Logit demand systems to capture the parsimony attained in characteristic space models in a product space specification of demand via exchangeability properties. Building on the discussion introduced in Section 4.3.1 they estimate a specification

$$\ln(s_{jt}/s_{0t}) = \alpha p_{jt} + x_{jt}\beta + f_j(\mathbf{s}_t, \mathbf{x}_t, \mathbf{p}_t) + \xi_{jt}. \quad (4.16)$$

It is convenient to rewrite $f_j(\cdot)$ as a function of $\{(s_{kt}, \mathbf{d}_{jkt})\}_{k \neq j}$, where $\mathbf{d}_{jkt} = \mathbf{x}_{jt} - \mathbf{x}_{kt}$, is the vector of distance in characteristics space (and price) to other products, as defined on page 36. We can think of this as a re-normalization that focuses on firm j and measures the distance of competitors from it. The key result is that in the Mixed Logit model $f_j(\cdot)$ is symmetric, or exchangeable, in its arguments (i.e., it depends on the states, but not their order) and does not depend on j .

⁴⁰Lee and Seo (2015) approximate $\sigma(\cdot)$ using Newton expansions for the purposes of proposing an alternative estimator to nested fixed point as discussed above.

A consequence of exchangeability, also discussed further in Gandhi and Houde (2020), is that $f_j(\cdot)$ can WLOG be represented as

$$f_j(\mathbf{s}_t, \mathbf{x}_t, \mathbf{p}_t) = g(EDF(\{(s_{kt}, \mathbf{d}_{jkt})\}_{k \neq j})),$$

where EDF denotes the empirical distribution function taken over the products in market t (specifically, a distribution over all products $k \neq j$ in market t). Based on this representation, they propose approximating f by using the first- and second-order set of empirical moments to approximate the EDF above. In principle, one could use higher moments as well. These moments are similar in spirit to the “within-group share” term in the Nested Logit model (4.10). These terms are endogenous, which couples closely with the IVs proposed in Gandhi and Houde (2020) that is based on a similar theoretical structure. They also use a flexible functional form for $g(\cdot)$, which they take to be a generalized additive model in each one of the moments used to approximate the EDF . There are other flexible approximations that may also work in practice. Finally, using the Implicit Function Theorem they show how these first stage estimates can be used to recover own- and cross-price elasticities.

To recover the distribution of heterogeneity, they recover ξ_{jt} as the residual from the demand equation (4.16). In the second step of their procedure they plug this residual into equation (4.15). This controls for the effect of the ξ_{jt} in the integral equation and allows them to estimate the distribution of heterogeneity, F , in a flexible parametric or non-parametric way using standard mixtures techniques. A key benefit of the approach even relative to the parametric Mixed Logit estimation of F discussed earlier is that it avoids the numerical complexity of demand inversion. This confers several benefits for the speed, reliability, and robustness of the estimator.

5 Supply

Having discussed ways to specify and estimate demand for differentiated products, we now turn to the supply side. Our focus is on pricing, but in principle the analysis below can apply to any continuous characteristic that can be flexibly adjusted.⁴¹ We have two goals in this section. First, we aim to show a few applications of the demand models

⁴¹For example, Fan (2013) looks at characteristics choice by newspaper, and how they change after a merger.

discussed earlier. Second, we present various ways in which the supply side has been modeled in the literature.

In some of the applications we will combine the demand model with a supply model to estimate the parameters of both demand and supply and test the supply model, as we saw in the motivating example discussed in Section 2. There are several ways to combine demand and supply. First, we could use equation (2.3) to recover marginal costs without assuming a parametric functional form for costs. Given demand estimates, a model of pricing and observed prices, we can back out the marginal costs that make the first-order condition for prices hold exactly for each observation in the data. Note, that the model will perfectly fit the data and therefore we cannot test it, unless we bring in additional information, such as information about marginal costs, and therefore markups.

Second, we can parameterize the marginal cost function and estimate its parameters, potentially jointly with the demand equation using the demand parameters estimated in a first stage. The advantage of estimating the marginal cost function is that it allows us to extrapolate to counterfactual situations not observed in the data. It will also allow us to test among models of competition, as we saw in Section 2, and estimate parameters associated with the pricing model.

5.1 The Workhorse Model of Horizontal Competition

The workhorse supply model in the study of differentiated-products industries is the static pricing model, described in Section 2.1.1.⁴² This model delivers a pricing equation given in equation (2.2), which can be used in a few different ways that we discuss below.

Berry et al. (1995) is a seminal paper in the development of equilibrium models of demand and supply in differentiated-products industries, and where much of the demand modeling discussed above was first developed. The authors were interested in understanding the impact of a voluntary export restraints placed on exports of automobiles from Japan to the United States. To study this question they developed and estimated a model of demand and supply in Berry et al. (1995) and applied it to this question in Berry et al. (1999).

The indirect utility in BLP is given by

$$u_{ijt} = x_{jt}\beta_i + \alpha \ln(y_i - p_{jt}) + \xi_{jt} + \varepsilon_{ijt}$$

⁴²We will refer to this model interchangeably as the workhorse model and Nash-Bertrand model.

where y_i is income and all other variables are as previously defined. Note, that income enters the indirect utility non-linearly and therefore will not cancel, as we discussed on page 18. For estimation they assume that income is distributed log normal and estimate the parameters of the distribution directly from income data. Consumer heterogeneity is described by

$$\beta_i^{(k)} = \beta_0^{(k)} + \beta_\nu^{(k)} \nu_i^{(k)} \quad \nu_i^{(k)} \sim N(0, 1).$$

Finally, the utility from the outside option is given by

$$u_{i0t} = \alpha \ln(y_i) + \varepsilon_{i0t}.$$

They estimate the model using 20 years of annual national data on the sales of automobiles in the United States. The model is estimated using both demand and supply-side moments. The demand-side IVs are the ones we described in Section 4.3.1. They also add supply-side moments as we described in Section 4.3.2.

The paper delivers two sets of important results. First, the paper demonstrates the importance of controlling for the endogeneity of price in the aggregate demand equation derived from discrete choice. Nowadays this is taken as obvious – not surprisingly since the importance of accounting for the endogeneity of price has been confirmed repeatedly in numerous industries – but at the time it was questioned as empirically relevant. For example, Table 3 of Berry et al. (1995) shows that a Logit demand model estimated without correcting for endogeneity of price yields a large number of inelastic demand curves (1494/2217=67.3%), which is inconsistent with static profit maximization. Once they instrument for price that number drops below 1%.

Second, the paper demonstrates the ability of the demand model to yield reasonable substitution patterns. BLP present these in Tables 5-7 . For example, in Table 7 the authors compare the diversion to the outside good, i.e., the fraction of consumers who substitute to the outside good in response to a price increase as a fraction of all those who substitute away from a product, implied by the Logit model and the more flexible random coefficients model. For the Logit model, as expected, the diversion is roughly 90% for all cars, which is roughly the share of the outside good. This number is high in absolute value, but maybe more important is that it is roughly constant: if the price of a BMW 735i increases, consumers are equally likely to divert to the no-purchase option as consumers of a Mazda 323. This seems unreasonable (and is totally driven by the Logit assumptions.) The outside option includes all the choices that do not involve purchasing a new car,

such as buying a used car, not replacing an existing car or delaying purchase. Intuitively, consumers who purchase a car are more likely to buy another car than switch to the no-purchase decision in response to a price increase, compared to the average consumer. Furthermore, consumers who purchased a lower priced car are more likely to switch to the no-purchase option in response to a price increase. In contrast to the Logit model, the full model can capture these effects: the overall substitution to the outside good is lower and the more expensive the car the lower the number. For example, the diversion for Mazda 323 is roughly 27% while for BMW 735i is 10%.

In large part, the paper is able to deliver statistically significant estimates of the variation in random coefficients because the authors impose the supply/pricing equation in estimation. As we noted earlier, the supply equation, jointly with a functional form for marginal costs that ensures that these costs will be non-negative, puts significant restrictions on the demand estimates. For example, demand estimates that imply inelastic demand also imply negative marginal cost under many models of pricing. A functional form that imposes (a reasonable assumption) that all marginal cost are positive would prevent this from happening. If we believe strongly in the supply model it is efficient to impose it in estimation. However, as we discussed in Section 2, and as we discuss below, a large motivation for estimating demand and supply models is precisely to test the supply model rather than assume it.

The paper abstracts from some elements of the car industry. For example, cars are a durable good. Yet the demand system is static: consumers are not forward-looking in the sense that they anticipate future needs, nor do they take into account whether they own a car and if so which one, when they make a purchase decision. On the supply side, the model assumes that manufacturers set a uniform price to consumers: dealers play no role and there is no price discrimination. In reality, we know that cars are sold through dealers and this market exhibits significant variation across consumers in the price they pay. Furthermore, the pricing model is static while in reality prices might reflect inventory considerations, generating brand loyalty or other dynamic effects. For many questions it is fine to abstract away from these issues, yet for some questions these issues might be quite important. Indeed, the paper inspired a large literature that relaxed some of these modeling assumptions on both the demand and supply side. We discuss some of the papers in this literature in the rest of this chapter and many others are discussed in other chapters in this Handbook.

5.2 Distinguishing Between Models of Competition

In this subsection we look to expand the supply, or pricing, models we consider. We will have a dual goal of coming up with potentially more general, or more flexible, models that would allow us to explain different patterns of pricing (and markups), as well as testing the workhorse, Nash-Bertrand model, and finding ways to distinguish it from alternative models. For the most part, the alternative models will focus on models that are more "collusive" and therefore tend to imply higher markups. We also use the discussion below as an opportunity to discuss several empirical implementations of the demand model.

A natural place to start thinking about expanding beyond the workhorse model is to consider testing of the model. Testing will generally follow the ideas presented in Section 2. As we saw there, one way to test the model is to informally compare predictions of prices and costs from different models to patterns we see in the data, even if only at an aggregated level. Nevo (2001) does precisely that. He studies pricing in the ready-to-eat cereal industry. The industry, at the time he studies it, was characterized by high concentration (the top 3 firms had approximately 75% share, and the top 6 approximately 90%), "high" price-cost margins (approximated to be around 45%), large advertising to sales ratios (roughly 13%) and numerous introductions of brands (67 new brands introduced by top 6 firms in 1980's). These facts were used to claim that this is a perfect example of an industry where firms collude on pricing but compete on advertising and brand introductions.⁴³ The paper asks if prices observed in this industry, and the margins that were approximated, are consistent with collusive pricing? Specifically, the paper notes that seemingly high margins can be due to product differentiation and multi-product pricing (of substitute products) and are not necessarily indicative of collusion.

To separate the effect of collusion Nevo estimates a brand-level demand system, he then computes price-cost margins implied by different pricing models and chooses the pricing model that cannot reject the approximated margins of 45%. He finds that the "high" margins are consistent with Nash-Bertrand pricing by multi-product firms and therefore one does not need to rely on collusion to explain the "high" margins. In other words, the data cannot reject the Nash-Bertrand model.

The demand model he uses follows equation (3.3). He estimates the model using scanner data for the top 25 brands of cereal. The data are aggregated to the MSA-quarter

⁴³For example, Scherer (1982) argues that "...the cereal industry's conduct fits well the model of price competition-avoiding, non-price competition-prone oligopoly" (p. 189)

level and he aggregates different SKUs to a brand. This results in 1,124 markets and 27,862 brand-quarter-MSA observations. He defines prices as total revenues divided by the number of servings, which he defines using the suggested servings size. Characteristics come from nutritional information (e.g., fat or sugar content), segment information used by the industry, and subjective information (e.g., he defines a “mushy” dummy variable). He estimates demand using the BLP algorithm and prices in other cities as IVs. Unlike BLP he estimates demand without imposing any supply-side moments, which has the advantage of yielding consistent estimates even if we are unsure about the supply model. He is able to do so because he has more markets than BLP and more variation in demographics across markets.

For the supply side he computes three models: single product, multi product and collusion, following the models discussed in Section 2.1.1. The markups implied by different supply and demand models are presented in Table 8 in the original paper. Using the random coefficients demand estimates he finds that the current ownership of the top 25 brands predicts an average margin of 42%, while joint ownership of these brands predicts a margin of 73% (with a confidence interval between 62% and 97%). By comparing these results to the approximation of the margins, which he estimates at 45%, he rejects the null of perfect collusion, or joint profit maximization of the top 25 brands, but cannot reject the null of Nash-Bertrand pricing.

Like BLP, this paper also raises several questions regarding the modeling assumptions. First, a common concern is whether demand is really discrete. In micro data, we often observe consumers purchasing more than one box of cereal, at times even different brands on the same trip. How do we reconcile this pattern with a discrete choice? One option is to think of the choice as happening at the time of consumption: at that point it seems more reasonable to assume a discrete choice. Thinking of the model this way ignores the two-step process, where the consumer first decides what to purchase at the store and then what to consume from the brands available at home, but with aggregate data it is unclear we can separate this process. In Section 6.1.1 we will discuss models that tackle the modeling and estimation of multiple choices.

The paper does not explicitly model retailer behavior despite using retailer prices to study manufacturer competition. This is consistent with retail margins going into the cost manufacturers pay to get the product on the shelf. This assumption, however, is not consistent with a strategic retailer who will change their margin in response to the

manufacturer pricing. In the next subsection, we will discuss how to add retailers to the model.

Finally, much of the price variation at the store level is coming from “sales”, or temporary price reductions. This creates an incentive for consumers to purchase the products when the prices are low and consume them later. Follow-up work by Hendel and Nevo (2006a,b), which we discuss in Section 6.2.1, follows up on this issue.

One of the common uses of the supply model is to simulate the effects of a proposed merger. The idea was discussed by Berry and Pakes (1993) and Werden and Froeb (1994), and implemented empirically by Nevo (2000a). The basic idea is as follows. Using pre-merger data one can estimate demand and recover marginal costs. The marginal cost can be recovered as in BLP, by parameterizing the cost function and estimating it jointly with demand. Alternatively, marginal cost can be recovered, without making any parametric assumptions on the cost function, by “inverting” the pricing equation (2.2) such that

$$\hat{m}c = p - \hat{\Omega}^{-1}q, \quad (5.1)$$

where $\hat{\Omega}^{-1}$ is computed using the demand estimates and q are observed quantities. One then uses these estimates to simulate the effect of the merger by changing the ownership structure defined in equation (2.1). In general, holding costs constant a merger between substitutes products will lead to higher prices, and a merger between complements will lead to lower prices. The real issue is not the direction of the effect, but the magnitude: if the products are closer substitutes the effect will be larger. The reason to estimate demand is to be able to quantify the effect. Furthermore, one could also use the model to simulate the impact of various efficiencies such as reductions in marginal costs, or improvements in qualities. This approach has found support in the academic literature and in practice.⁴⁴

Simulating the effects of mergers can be the object of interest, but can also serve as a way to test the model. An early attempt at this was provided by Peters (2006) who used comparisons between predicted and actual outcomes of airlines mergers to test the demand model. Miller and Weinberg (2017) study the beer industry and use a joint venture, which they treat as a merger, to test the Nash-Bertrand model. Their study is motivated by a desire to understand the price effects after the 2008 Miller-Coors joint venture (JV). The JV significantly increased concentration in the industry. After it, ABI (the producer

⁴⁴For a discussion of the use in policy see Griffith and Nevo (2019), and for discussion of the use of merger simulation in a specific merger see Bayot et al. (2018).

of Budweiser) had a 37% market share, Miller-Coors had 29%, Modelo (the producer of Corona) had 9%, and Heineken had 5%.

They document, in Figure 1 (and Tables 2 and 3) of their paper, that post-JV (i) the price of Miller-Coors products increased; (2) the price of ABI increased almost one-to-one with the Miller-Coors price; and (3) the prices of Modelo and Heineken did not increase. The direction of the first fact is not surprising, since, as we discussed above, this is what a standard Nash-Bertrand model would predict. The magnitude might be higher than expected, but we cannot measure that without knowing the cross-price elasticities between the products. The fact that ABI prices increase should also not be surprising since prices are strategic complements in a pricing game for many (but not all) demand models. In these cases, the increase in the price of Miller-Coors would lead to an increase in the ABI price. What is surprising is that the ABI price increase is similar in magnitude to the Miller-Coors price increase. The paper sets to see if the demand estimates can reconcile these increases with what a competitive pricing model would predict.

The demand model they estimate is a random coefficients Nested Logit model, where all the products are in one nest and the outside good is in another. They estimate the model using monthly (or quarterly) scanner data for 13 brands in 39 different distinct regions, and the BLP nested fixed point algorithm discussed above. For IVs they use the distance between the brewery and the region, an indicator variable equal to one for ABI and Miller-Coors products after the merger, the number of products, mean income (by region) interacted with product characteristics. Their estimates (Table 4 in the paper) find statistically significant heterogeneity in preferences, but economically fairly small. The implied elasticities (presented in Table 5) show a slight variation from the pattern implied by the model with no heterogeneity: substitution between the inside goods is roughly proportional to share.

To estimate the supply side they specify marginal cost as a function of distance from the brewery, an indicator equal to 1 one for Miller-Coors products post-merger, and product region and period fixed effects. They use the model of Section 2.1.1 but slightly modify the definition of the ownership, given in equation (2.1). Specifically, they assume that the (j, k) element equals κ if products j and k are sold by ABI and Miller-Coors post-merger. This generates Nash-Bertrand competition in the post-merger periods if $\kappa = 0$ and joint profit maximization for ABI and Miller-Coors if $\kappa = 1$. Putting this together with equation

(4.12) yields the estimating equation

$$\mathbf{p}_t = \mathbf{w}_t\boldsymbol{\gamma} + \Omega^{-1}(\kappa)\mathbf{q}(\mathbf{p}_t) + \boldsymbol{\omega}_t,$$

where variables are defined as in (4.12). They estimate the cost parameters and κ using GMM and assuming that ω_{jt} is mean independent of w_j and an indicator equal to one for ABI and Miller-Coors products post merger.⁴⁵

They estimate that κ varies between 0.25 and 0.34, depending on the specification they use. All the specifications reject the null hypothesis that $\kappa = 0$, which is Nash-Bertrand pricing post merger. Interestingly, the model also rejects joint profit maximizing pricing, between ABI and Miller-Coors post merger. In other words, pricing is consistent with some increased coordination post merger but not joint profit maximization.

Ciliberto and Williams (2014) use a similar idea to study multi-market contact in the U.S. airline industry. They modify equation (2.1) by making the terms H_{jr} a function of a variable that measures the degree of multi-market contact between the airlines that produce products j and r . They conclude that airlines with a higher degree of multi market contact almost perfectly collude.

It is tempting to try to interpret intermediate values of κ . For example, one interpretation, which was popular for a while in the analysis of competition in homogeneous goods industries, views κ as a "conduct parameter" that captures beliefs about the equilibrium being played (Bresnahan, 1989). A slightly different interpretation views κ as an "as if" parameter. For example, if $\kappa = 0.5$ in a homogeneous good industry, the industry would be seen as being as competitive as an industry with 2 symmetric firms playing Cournot. Similar "as if" potential interpretations exist for industries with differentiated products. For example, building on Nevo (1998), Black et al. (2004) propose an interpretation of κ as a measure of cross ownership.⁴⁶

These attempts to interpret intermediate values of κ have fallen mostly out of favor for several reasons. For example, Corts (1999) notes that if the true model implies variation in κ over time within each regime, the "as if" interpretation is problematic. Indeed, he shows via simulations that the estimated κ need not even be a good indicator of the relative competitiveness of industries. His main complaint is not about the idea of having

⁴⁵Note that prices appear on both sides of the estimating equation, and therefore we need an (additional) IV, in this case an indicator equal to one for ABI and Miller-Coors products post merger.

⁴⁶Backus et al. (2021) build on these ideas, and use data from the ready-to-eat cereal industry, to test recent claims that common ownership facilitates collusion.

a parameter κ but to the a-theoretical restrictions put on it. He calls for a specification of a (structural) model of collusion in order to justify restrictions on κ . Miller and Weinberg (2017) are partly immune to this critique because of their focus on testing whether κ is statistically different than zero as a test for post-merger Nash–Bertrand competition (since under the null the model is well specified).

Most of the above discussion of testing and distinguishing between models was somewhat informal. However, the ideas are can be formalized. Bresnahan (1982) and Lau (1982) discuss the use of rotation of the demand curve as an IV to estimate the model of competition in homogeneous goods industries. Berry and Haile (2014) show that the basic idea generalizes to models that study competition in differentiated products industries. Specifically they show that the model of competition can be identified using the supply-side conditional moments $E(\omega_{jt}|\mathbf{Z}_t) = 0$, as defined in (4.12). The intuition behind this is what we saw in the applications above: different models of competition have implications for patterns we observe in the data. The patterns can be either average markups (Nevo, 2001), variation in prices after a merger (Miller and Weinberg, 2017), or co-variation in prices and multi-market contact across markets Ciliberto and Williams (2014). The moment condition is a way to summarize this variation.

Casting the problem as a moment condition allows Berry and Haile (2014) to formally discuss identification. It also has the advantage of generalizing the idea of using a specific event, for example, a merger, to using variation like that discussed in Section 4.2.2, to distinguish models of competition. This raises the question of what IVs should be used. Active areas of research are what IVs have power to distinguish models of competition (Duarte et al., 2021) and somewhat related is what is the efficient ways to use these IVs (Backus et al., 2021).

5.3 Adding Retailers Into the Mix

One of the common complaints about some of the papers presented in the previous subsection is that they are are concerned with competition between manufacturers yet use retail data without explicitly modeling the retailers. This practice can be justified if we treat the retailers as passive and therefore part of the manufacturer costs of getting the products to market. In this subsection we discuss papers that allow retailers to make strategic decisions in response to price changes by the manufacturers.

Villas-Boas (2007) asks what model best describes the relationship between manufacturers and retailers in the yogurt industry. To answer this question she takes the idea of

using demand estimates jointly with a pricing model to recover cost, and adds to it an additional layer: her pricing models account for both retailer and manufacturer/wholesaler behavior. She computes markups/costs under different models. She then chooses the model that best describes the data.

To estimate demand she uses weekly scanner data for 43 products, produced by 5 manufacturers and sold by 3 retailers. She observes retail prices, advertising, aggregate quantity (by product-retailer), and product characteristics, but does not observe wholesale prices or contracts. She estimates a random coefficients Logit Model, where the choice is store-brand. In other words, consumers choose not just the brand but also where to purchase it.

She considers six supply models. In Model 1, manufacturers first set (linear) wholesale prices and then retailers, taking these wholesale prices as given, set retail prices. In each stage there is Nash-Bertrand pricing. This results in double marginalization. Formally, retailer r in market t maximizes profits given by

$$\pi_{rt} = \sum_{j \in \mathcal{J}_{rt}} [p_{jt} - p_{jt}^{\omega} - c_{jt}^r] q_{jt}(\mathbf{p}_t).$$

where p_{jt}^{ω} is the wholesale price paid by the retailer, c_{jt}^r is the retailer's marginal cost and \mathcal{J}_{rt} is the set of products sold by the retailer. In her setting each "product" j is a brand-retailer combination and therefore this setting allows for different wholesale and retail prices to be charged for the same physical product. Rearranging the first-order conditions we can write

$$\mathbf{p}_t - \mathbf{p}_t^{\omega} - \mathbf{c}_t^r = (\Omega^r)^{-1} \mathbf{q}(\mathbf{p}_t),$$

where Ω^r is a matrix with elements given by $\Omega_{jk}^r = -\partial q_k / \partial p_j \cdot H_{jk}^r$, where j indexes rows and k columns and H_{jk}^r is the retailer ownership structure, namely $H_{jk}^r = 1$ if both j and k are sold by r .

The manufacturers' problem is to maximize profits given by

$$\pi_{\omega t} = \sum_{j \in \mathcal{J}_{\omega t}} [p_{jt}^{\omega} - c_{jt}^{\omega}] q_{jt}(\mathbf{p}_t)$$

where c_{jt}^{ω} is the manufacturer's marginal cost and $\mathcal{J}_{\omega t}$ is the set of products produced by manufacturer ω . Rearranging yields the following first-order conditions

$$\mathbf{p}_t^{\omega} - \mathbf{c}_t^{\omega} = (\Omega^{\omega})^{-1} \mathbf{q}(\mathbf{p}_t),$$

where Ω^ω is a matrix with elements given by $\Omega_{jk}^\omega = -\partial q_k / \partial p_j^\omega \cdot H_{jk}^\omega$, where j indexes rows and k columns and H_{jk}^ω is the manufacturer ownership structure. Note, that $\partial q_k / \partial p_j^\omega$ can be computed using the derivatives of the retail prices with respect to wholesale prices (by totally differentiating the retailers first-order conditions with respect to wholesale prices) and the derivatives of quantity with respect to retail prices.

In Model 2 the national brands are as in Model 1. However, the private labels are treated as a vertically integrated (the “manufacturer” sets the retail price for them).

In Model 3 she explores non-linear pricing. She consider two cases.⁴⁷ In case 1, manufacturers set the wholesale price equal to cost, $p_t^\omega = c_t^\omega$ and set fees F , aimed at extracting the retailers’ profits. In this case, the retail price is

$$\mathbf{p}_t - \mathbf{c}_t^\omega - \mathbf{c}_t^r = (\Omega^r)^{-1} \mathbf{q}(\mathbf{p}_t)$$

In case 2, retailers’ margins are set to zero, $p_{jt} = p_{jt}^\omega + c_{jt}^r$ and the retailers set a fee F to recover profits. Prices are set to maximize downstream profits for the manufacturers

$$\mathbf{p}_t - \mathbf{c}_t^\omega - \mathbf{c}_t^r = (\Omega^{\omega r})^{-1} \mathbf{q}(\mathbf{p}_t)$$

where $\Omega^{\omega r}$ is a matrix with elements given by $\Omega_{jk}^{\omega r} = -\partial q_k / \partial p_j \cdot H_{jk}^\omega$. The difference between the two cases is in the “ownership” structure used. In case 1 the pricing internalizes cross-price effects across brands *within* a store, while in case 2 the pricing internalizes cross-price effects within a brand *across* stores.

Models 4-6 allow for coordination. Model 4 is like Model 1 but with manufacturer-level (perfect) collusion. Model 5 is like Model 1 but retail-level (perfect) collusion. Finally, Model 6 offers a version of a vertically integrated monopolist.

Estimation of demand follows the standard approach, using as IVs cost data multiplied by a product fixed effect (i.e., the costs are allowed to impact each product differently). To choose between the supply models she uses two approaches. In the first she regresses

$$\mathbf{p}_t = \mathbf{w}_t \gamma + SIPC M_r(\Omega) \lambda_r + SIPC M_\omega(\Omega) \lambda_\omega + \boldsymbol{\omega}_t,$$

where \mathbf{w} are cost variables, γ is a vector of coefficients, $SIPC M_r(\Omega)$ and $SIPC M_\omega(\Omega)$ retail and manufacturer price-cost margins, respectively, implied by the different scenarios

⁴⁷See Bonnet and Dubois (2010) for the assumptions required to justify the first-order conditions that follow as coming from two-part tariff contracts.

above. She then tests if $\hat{\lambda}_r$ and $\hat{\lambda}_\omega$ are statistically different from 1.⁴⁸ The logic is that in a well specified model the observed prices should equal marginal cost plus the wholesale and retail margins, i.e., the coefficients on markups should equal 1.⁴⁹ Next, she uses a statistical test of non-nested alternatives, as in Bresnahan (1987) and Gasmi et al. (1992), using the procedure in Smith (1992).⁵⁰ Based on these she concludes that models that assume zero wholesale margins and in which retailers have pricing decisions best fit the data. She then discusses several contracts that are consistent with these outcomes.⁵¹

The paper shows that the logic of backing out markups and costs, presented in Section 2 can be extended to another (vertical) level of decisions. Indeed, the pricing equations she proposes are direct extensions of the pricing equation we discussed in the previous two subsections.

5.4 Models of Bargaining

Up to this point the supply models we focused on dealt with situations where one side of the market makes a take-it-or-leave-it offer. In the basic model, it was the manufacturers making a take-it-or-leave-it offer to consumers or to retailers. In the previous subsection we allowed retailers to make the offer to the manufacturers. In many real-world situations neither of these fits what actually happens because, for example, the parties, say a manufacturer and a retailer, negotiate the final outcome rather than one side dictating the terms of trade and the other side simply accepting or rejecting the offer. Even if there is no explicit negotiation, a bargaining model seems like a good way to model pricing that is between either extreme of prices set by upstream or downstream firms.

To fix ideas, we focus on a specific example, the study by Crawford and Yurukoglu (2012) of bargaining in TV markets. Crawford and Yurukoglu study the impact on consumer welfare of (un)bundling of TV channels offered in cable bundles. Generally cable companies offer consumers a bundle of channels and consumers cannot subscribe

⁴⁸Note that like our discussion of the supply model in Miller and Weinberg (2017), the parameters entering the pricing equation might require IVs to identify them because of the fact that the markup terms $SIPCM_r$ and $SIPCM_\omega$ might also be functions of prices and shares and therefore correlated with the econometric error ω_t .

⁴⁹Pakes (2017) conducts a similar exercise. Specifically, he uses demand estimates from Wollmann (2018) to construct the markup term. He then regresses the estimated markup on IVs and regresses the observed price on the observed cost determinants and this predicted markup. He finds that the estimated coefficient is not statistically different than 1.

⁵⁰See Rivers and Vuong (2002) and Duarte et al. (2021) for additional discussion of non-nested testing.

⁵¹See Bonnet and Dubois (2010) for follow up work that examines these issues.

to channels à la carte. Some have suggested that consumers are hurt by this arrangement because they have to pay for channels that they do not want. Crawford and Yurukoglu (2012) ask: what are the (equilibrium) welfare effects of unbundling? Holding channel prices constant, consumers will be better off with unbundling because they could purchase “skinnier” bundles. However, Crawford and Yurukoglu point out that this argument ignores equilibrium effects. Therefore, to answer the question of whether consumers are better or worse off with unbundling, they need to model the effect of unbundling on input prices, namely the prices paid to the content providers. To do so they develop an empirical bargaining model based on the theoretical model of Horn and Wolinsky (1988a,b).

They model the negotiation between each conglomerate of channels (e.g., ABC Disney, which owns ESPN, ESPN2, Disney channel and other channels) and the operator/distributor (e.g. Comcast), as a bilateral bargaining problem over linear (input) prices for that pair. The Nash bargaining solution is the set of input prices that maximize the weighted product of the values to both parties from agreement (as a function of that price) relative to the values without agreement. In general, the Nash bargaining solutions are interdependent as the value from one agreement depends on what happens in the other agreements. For example, if negotiations break down with one conglomerate the value of another conglomerate to the distributor might increase. To make progress (Horn and Wolinsky, 1988a,b), make a Nash equilibrium like assumption: they assume that each negotiating pair takes the outcome of other negotiations as given.⁵² This setup is often called “Nash-in-Nash” bargaining. As Crawford and Yurukoglu point out, this is a strong assumption, since the members of each pair also participate in other negotiations and realize that whether an agreement is reached, and what is that agreement, will impact other negotiations. Furthermore, because the negotiating parties participate in other negotiations they will have asymmetric information. Nevertheless, in order to make progress much of the empirical work has relied on this assumption.⁵³

Formally, let p_{rj}^{ω} be the wholesale price, or input cost, paid by distributor r for channel j , which is owned by conglomerate f . Note that the distributors in this setting are equivalent to downstream retailers and the conglomerates are the equivalent of upstream

⁵²In bilateral negotiation, Rubinstein (1982) and Binmore et al. (1986) show that the Nash bargaining solution corresponds to the unique subgame perfect equilibrium of an alternating offers non-cooperative game. Collard-Wexler et al. (2019) provide conditions such that the Horn and Wolinsky solution is the same as the unique perfect Bayesian equilibrium with passive beliefs of a specific simultaneous alternating offers game with multiple parties.

⁵³See Lee and Fong (2013), Ho and Lee (2019), Ghili (ming) and Liebman (2018) for alternatives.

manufacturers. Let \mathbf{p}^ω denote the vector of all wholesale prices, and let $\mathbf{p}_{r,f}^\omega$ be the vector of wholesale prices paid for the channels of conglomerate f . The conglomerate and the distributor negotiate over the vector of wholesale prices, $\mathbf{p}_{r,f}^\omega$. If negotiations break down the distributor will not have access to any of the conglomerate channels. The Nash bargaining solution determines the prices $\mathbf{p}_{r,f}^\omega$, taking as given all other prices $\mathbf{p}_{-r,f}^\omega$, to maximize distributor r 's and conglomerate f 's Nash Product, defined by

$$NP_{rf}(\mathbf{p}_{rf}^\omega; \mathbf{p}_{-rf}^\omega) = [\pi_r(\mathbf{p}_{rf}^\omega; \mathbf{p}_{-rf}^\omega) - \pi_r(\infty; \mathbf{p}_{-rf}^\omega)]^{\zeta_{rf}} [\pi_f(\mathbf{p}_{rf}^\omega; \mathbf{p}_{-rf}^\omega) - \pi_f(\infty; \mathbf{p}_{-rf}^\omega)]^{(1-\zeta_{rf})}, \quad (5.2)$$

where ζ_{rf} is a parameter that measures the (relative) bargaining power of r when bargaining with f , and π_r and π_f are the profit functions of r and f , respectively, when an agreement is reached and when it is not (denoted by $\mathbf{p}_{r,f}^\omega = \infty$). Note, that the infinite price in the case of disagreement reflects that the distributor will not have access to any of the conglomerate channels. These profit functions are determined endogenously by parts of the model explained below.

Two key determinants of the bargaining outcome are the bargaining power, captured by ζ_{rf} , and the bargaining leverage, which is the loss from disagreeing relative to reaching an agreement. The bargaining parameter varies between zero, where f has all the bargaining power and the solution is equivalent to Nash-Bertrand pricing by the upstream providers, and one, where r has all the bargaining power and can impose its will (subject to a participation constraint by f). Values between the two extremes allow for different relative bargaining power.⁵⁴

The leverage of the parties impacts the solution in a similar way: everything else equal the bargaining solution is closer to the optimal solution of the party with the greater leverage, namely the party with less to lose from a disagreement. For example, if distributor r is negotiating with two different conglomerates f and f' and the only difference is that

$$\pi_f(\mathbf{p}_{rf}^\omega; \mathbf{p}_{-rf}^\omega) - \pi_f(\infty; \mathbf{p}_{-rf}^\omega) > \pi_{f'}(\mathbf{p}_{rf'}^\omega; \mathbf{p}_{-rf'}^\omega) - \pi_{f'}(\infty; \mathbf{p}_{-rf'}^\omega).$$

Namely, that relative to f' , f has more to gain from an agreement, or more to lose from a disagreement, then f' has more leverage and will obtain an outcome that is more favorable from its perspective.

⁵⁴To formally see these claims, note that the maximization problem is equivalent to maximizing the log of the Nash product. Taking the first-order condition of the log of the Nash product and slightly rearranging we see that the resulting first-order condition is just a weighted average of the maximization problems of r and f , with the weights being a product of the relative bargaining leverage and bargaining power.

To compute the profit functions of both the conglomerate and the distributors Crawford and Yurukoglu estimate a viewership model and bundle (subscription) choice model. They then back out the implied input cost from the bundle pricing equation and use it to fit a parametric cost function for the distributors. The details of their modeling (and estimation) are a bit different from, for example, BLP, but the principle is the same: the pricing equation and demand estimates are used to back out input costs or estimate an input cost function. These input costs can then be used, the same way prices were used in the basic supply model, to either recover costs (here these are upstream costs) or parameters of the supply model (here bargaining parameters).

In their application, Crawford and Yurukoglu assume that upstream marginal costs are zero and therefore they estimate the bargaining parameters. They do so by choosing the bargaining parameters, ζ_{rf} that minimize the difference between the programming costs they recovered from the pricing and demand equations and those predicted by the bargaining model. One way to implement this idea is to write

$$\hat{p}_{rj}^{\omega} = p_{rj}^{\omega*}(\zeta_{rf(j)}) + \epsilon_{rj} \quad (5.3)$$

where \hat{p}_{rj}^{ω} is the wholesale price backed out from the pricing equation, $p_{rj}^{\omega*}(\zeta_{rf(j)})$ is the wholesale price that maximizes the Nash product defined in (5.2) and ϵ_{rj} is measurement error, or an "add-on" error term. The parameters ζ can be estimated using, for example, non-linear least squares.

Note, that the bargaining parameters vary by conglomerate and not channel and therefore in principle are of lower dimension than the wholesale prices. In practice, Crawford and Yurukoglu (2012) do not have enough variation to meaningfully estimate channel-level wholesale prices and instead estimate these at the conglomerate-level. This is of little economic significance because the way they set up the bargaining model all that matters is the total payment to each conglomerate. Furthermore, for computation reasons they solve the model for a market with "typical" distributors. The bottom line is that they have the same number of parameters as observations in the above equation and therefore can compute the set of bargaining parameters that set the backed out input cost exactly equal to those predicted by the bargaining model. This directly parallels the non-parametric approach to recovering cost discussed in the Introduction to this section. In this approach some view the bargaining parameters as the "error terms."

In other settings (Grennan, 2013; Bagwell et al., 2020), the number of observations in (5.3) is larger than the number of bargaining parameters and then those are estimated,

for example, by non-linear least squares.⁵⁵ In principle, the bargaining parameters can be estimated jointly with the pricing and demand parameters, but computationally it is easier to estimate them separately.

Using the estimated parameters Crawford and Yurukoglu are able to simulate the model both with bundling, as in the data, and when the consumers are offered channels à la carte. They find that equilibrium effects are significant: the counterfactuals suggest that à la carte offering of channels will increase input prices enough to offset consumer benefits from being able to choose only the channels they value most. Thus, unbundling will decrease welfare.

It is possible to set up, at least for some bargaining problems, a “structural” error term that follows a bit more closely the logic of the model discussed in Sections 5.1-5.3. This approach especially makes sense when the upstream costs are unknown and need to be estimated: it seems natural to assume that the error term consists of unobserved variation in marginal costs. Gowrisankaran et al. (2015) do precisely that.⁵⁶ They study the effects of mergers when prices are negotiated using data from negotiations between hospitals and health insurance companies, or managed care organizations (MCOs), to demonstrate the effect. Following Town and Vistnes (2001) and Capps et al. (2003) they set up the bargaining problems as follows. Let r denote the MCO, and f denote a system of hospitals that includes hospitals $j \in \mathcal{J}_f$. The negotiation is over the price that the MCO will pay the hospital if one of its enrollees receives care in that hospital.

In this setting the MCO is the equivalent of the retailer or distributor, and the hospital system is the equivalent of the manufacturer or conglomerate. The Nash product for this problem follows equation (5.2), with a small difference that the profit functions, and therefore the Nash product, are a function of wholesale prices and the “network” (i.e., which MCO-hospital system pairs that reach an agreement).⁵⁷ Specifically, hospital system, f profit from an agreement relative to disagreement is given by

$$\sum_{j \in \mathcal{J}_f} q_{rj}(\mathcal{N}, \mathbf{p}^\omega) [p_{rj}^\omega - mc_{rj}], \quad (5.4)$$

⁵⁵For consistency of the estimates, generally, the error term in (5.3) cannot appear inside $\mathbf{p}^{\omega*}(\cdot)$ and therefore needs to be an “add-on” error term that is usually interpreted as “measurement error”.

⁵⁶Grennan (2013) also estimates costs parameters, but his econometric error term is in the bargaining parameter.

⁵⁷This is not a meaningful distinction because an infinite wholesale price can denote disagreement. The main reason for the change in notation is to be more explicit about networks, which are of greater importance in health care where not all MCOs contract with all hospital systems.

where p_{rj}^ω is the price r pays f if one of its enrollees is treated in hospital j , \mathbf{p}^ω is the collection of all prices, \mathcal{N} is a description of the network (i.e., which hospitals have contracts), $q_{rj}(\mathcal{N}, \mathbf{p}^\omega)$ are the patients that go to hospital j given the network and the vector of prices, mc_{rj} is the marginal cost of treating those patients (which are assumed (i) constant in quantity; and (ii) to vary by MCO to allow for different composition of enrollees.) This formulation assumes that patients do not go out of network and will not switch insurance providers in the case of disagreement.

On the MCO side, the value of an agreement is given by $\pi_r(\mathcal{N}_r, \mathbf{p}^\omega) - \pi_r(\mathcal{N}_r \setminus \mathcal{J}_f, \mathbf{p}^\omega)$. $\pi_r(\cdot)$ is the value to the MCO of having a network (with and without system f). Note, that as before, a breakdown in negotiation means the whole system leaves the network. The paper builds $\pi_r(\cdot)$ from micro-foundation based on consumers' willingness-to-pay for an insurance product that has a wider network.

They show that one can write the first-order conditions of the bargaining problem as

$$(1 - \zeta_{rf}) \frac{q_{rj} + \sum_{k \in \mathcal{J}_f} \frac{\partial q_{rk}}{\partial p_{rj}^\omega} [p_{rk}^\omega - mc_{rk}]}{\sum_{k \in \mathcal{J}_f} q_{rk} [p_{rk}^\omega - mc_{rk}]} = - \zeta_{rf} \frac{\overbrace{\frac{\partial \pi_r(\mathcal{N}_r, \mathbf{p}^\omega)}{\partial p_{rj}^\omega}}^A}{\underbrace{\pi_r(\mathcal{N}_r, \mathbf{p}^\omega) - \pi_r(\mathcal{N}_r \setminus \mathcal{J}_f, \mathbf{p}^\omega)}_B}. \quad (5.5)$$

The assumption of constant marginal costs implies that the first-order conditions in equation (5.5) are separable across MCOs and therefore one can rearrange the joint system of $\#(\mathcal{J}_f)$ first-order conditions from (5.5) to write

$$\mathbf{q} + \Omega(\mathbf{p} - \mathbf{mc}) = -\Lambda(\mathbf{p} - \mathbf{mc}) \quad (5.6)$$

where Ω and Λ are both $\#(\mathcal{J}_f) \times \#(\mathcal{J}_f)$ size matrices, with elements $\Omega_{jk} = \frac{\partial q_{rk}}{\partial p_{rj}^\omega}$ and $\Lambda_{jk} = \frac{\zeta_{rf}}{1 - \zeta_{rf}} \frac{A}{B} q_{rk}$. Solving for the equilibrium prices yields

$$\mathbf{p} = \mathbf{mc} - (\Omega + \Lambda)^{-1} \mathbf{q}, \quad (5.7)$$

where \mathbf{p} , \mathbf{mc} and \mathbf{q} denote the price, marginal cost and quantity vectors respectively for an MCO r across the different hospitals. Equation (5.7), which characterizes the equilibrium prices, has a form almost identical to standard pricing games, but differs in the inclusion of Λ . One case where $\Lambda = 0$ – and hence there is differentiated-products Nash-

Bertrand pricing with individual prices for each MCO – is where hospitals have all the bargaining weight, $\zeta_{rf} = 0, \forall f$.

Importantly, equation (5.7) shows that, as with static differentiated-products pricing models, we can back out implied marginal costs for the bargaining model as a closed-form function of prices, quantities and derivatives, given MCO and patient incentives. Furthermore, one can combine equation (5.7) with an assumption about cost, as in Section 4.3.2,

$$mC_{rj} = w_{rj}\gamma + \omega_{rj}$$

to form a basis for estimation where ω_{rj} is the structural error term and can be used to form a GMM objective function, as in equation (4.12).

The above discussion makes it clear that bargaining models can be specified in a variety of ways that impact formulation as well as identification and estimation. Our discussion is just an introduction to the topic, which is further explored by Lee et al. (2021) in this Handbook.

6 Extensions of the Demand Model

In this section we discuss some of the extensions of the discrete choice demand model discussed in Section 3. We start by outlining some extensions to the static model and then discuss models of dynamic demand.

6.1 Extensions to the Static Demand Model

A reasonably large literature has explored several extensions to the basic discrete choice models discussed in Section 3. The motivation for these extensions varies. Some extensions are motivated by a desire to better marry the model with patterns observed in micro data, and potentially explore the biases that would arise if we ignore these differences. Other extensions are needed to address specific questions that the basic model could not address. In many cases, the extensions are estimated using individual data, and therefore might not generalize to the various applications previously discussed.

Here we focus on two such extensions: multiple goods and a more general characteristics model. There are other extensions that have been considered in the literature. For example, Dubin and McFadden (1984) and Hanemann (1984) propose a model where consumers make a discrete choice, say choose an appliance, followed by a continuous

choice, for example, how much to use that appliance. The two decisions are obviously interlinked. These ideas have been used more recently in looking, for example, at store choice (Smith, 2004; Thomassen et al., 2017) and in choice of internet service plan and usage (Malone et al., 2020; Nevo et al., 2016).

Other areas that have received considerable attention are models where consumers do not know the characteristics of the products and have to learn them through search. See Honka et al. (2019) for a survey of this literature.

6.1.1 Multiple goods

A common complaint about discrete choice models is that actual purchases, when they are observed, are not of a single product or a single unit. For example, when studying the purchase of cereal we might observe consumers buying, on a single shopping trip, several boxes of the same cereal brand or buying several brands. One way to rationalize the multiple choices is to assume a single purchase instance, which is what we observe, is an aggregation over several consumption instances, which is what we are modeling. For example, a consumer shopping in a store might be buying for several household members for a week. If we assume that there are 3 household members, each making a daily consumption choice then we can view the single observed purchase as aggregation over 21 consumption choice instances. This can help rationalize purchase of multiple units, as well as different brands, and is probably sufficient when working with aggregate data where the individual choices are not observed anyway and we are already aggregating across consumers. However, this explanation is somewhat unappealing when working with individual-level data, in part because it assumes the choices across days are independent. Therefore, a somewhat better alternative is to more explicitly build the purchase model from underlying choices.

One way to model multiple choices is to redefine what we mean by an option. For example, assume that there are two brands, A and B, and consumers can buy no more than 2 units of each. This gives 9 options, ranging from not buying either brand to buying 2 of each brand. The utility from each option could in principle, be additive, in the utility from each brand separately or have interactions. For instance if buying one unit of each brand there could be an interaction term, which can be positive or negative, so that the utility is not just the sum of the utilities of buying each brand.

Gentzkow (2007) takes this approach when studying consumers' choice between print and online newspapers. He allows for purchase of more than one option and estimates

an interaction term in the utility function (to understand if print and online newspapers are complements of substitutes.) In his setting, consumers choose between the printed version of a newspaper, the online version, both, or neither. Thus, the choice is a discrete choice between bundles. He further assumes that each bundle receives an i.i.d. shock, as in the standard discrete choice model, and therefore he can estimate the model using standard methods. Note, that this assumption requires that any correlation between options, which include the same products, is captured through the rest of the utility terms. Furthermore, this approach is feasible when the number of choices is small, since the number of bundles increases exponentially with the number of products. For example, with $J = 20$ different options and even if the consumer can only choose at most one of each there are $2^{20} = 1,048,576$ different bundles available.

Another approach, proposed by Hendel (1999), and later also used by Dubé (2004) builds the demand explicitly from underlying tasks. Specifically, Hendel observes firms simultaneously buying several brands of computers and several units of each brand. To model this, he assumes that the firm has several tasks to do. For each task there is an optimal choice, but the observed purchases are an aggregation over several tasks. He does not allow for interaction in the utility from the different choices. He explains the purchase of several units, of the same computer, by a decreasing marginal utility from quantity, hence there is interaction in this dimension.

Fan (2013) studies mergers in the newspaper industry. She allows consumers to buy more than a single newspaper (but at most one copy of each) and introduces a parameter that measures the decrease in utility from a newspaper if it is bought "second". Therefore, like Gentzkow (2007) she allows for an interaction in utility, but at the same time, like Hendel (1999), does not treat the bundle as an independent option.

Nevo et al. (2005) study the decision of libraries to subscribe a subset of the 150 or so to Economics and Business journals that they observe. The libraries in their data subscribe to some subset of these 150 journals, but the subsets are not nested. If they were nested, we could model the choice of a bundle as a choice of how many journals to purchase. Instead, they model the choice problem as ranking journals by an index like that standard index in discrete choice models, for example the one given in equation (3.3). Unlike the standard discrete choice problem where the decision maker only chooses the top option, here the decision maker subscribes to journals following this ranking until a (budget or other) constraint is met. The utility from the journals does not interact (i.e., the utility does not depend on what other products are in the bundle), but the interaction is through

a (budget) constraint. They show how one could estimate this model using library-level subscriptions.

6.1.2 General Characteristics Demand Models

In Section 3.1, when we discussed the various demand models, we separated models in product space from models in characteristics space. For models in characteristics space we focused on discrete choice models. The idea of using characteristics to reduce the dimension of the estimation problem can be useful more generally. This turns out to be especially helpful when trying to estimate demand for more than a single product category. As such, this model offers an additional way to deal with multiple purchases.

Dubois et al. (2014) study such a problem.⁵⁸ They note, looking at household-level purchase data from France, the United Kingdom and the United States, differences across countries in the choices households make and in the prices and product offerings they face. These differences amount to large difference in the nutritional intake. For example, the average French consumer purchases roughly 200 calories less a day than the average American consumer. They ask the extent to which cross-country differences in purchases are attributable to differences in prices and the characteristics of products (the economic environment). To address this question they develop a model in characteristics space, based on Gorman (1980) and Lancaster (1966), that does not assume discrete choice. Their motivation is to use the richness of their data, which include disaggregated purchases, while still looking at the choice of a food "bundle" rather than narrowly defined products (e.g., soft drinks).

Specifically, a consumer i , with demographics D_i , chooses from J products, where product j is characterized by K characteristics $\{a_{j1}, \dots, a_{jK}\}$ and K is much smaller than J . These characteristics will in principle include both observed characteristics, like calories and protein in their example, and unobserved characteristics, which we previously denoted by ξ . The utility is given by $U(q_0, \mathbf{x}, \mathbf{q}; D_i)$ where q_0 is the amount of the numeraire good consumed, \mathbf{x} is a $K \times 1$ vector of characteristics of food consumed by the consumer across all the products and \mathbf{q} is a $J \times 1$ vector of the quantities purchased of all food products. Define the $J \times K$ matrix $\mathbf{A} \equiv \{a_{jk}\}_{j=1, \dots, J, k=1, \dots, K}$. This matrix will be

⁵⁸See also Pinkse et al. (2002) who study competition between gas stations, and Pinkse and Slade (2004) who study the beer market. In both cases they model demand using a system in products space, and model the substitution matrix as a function of either physical distance or distance in attribute space. The resulting system is a statistical mapping between price and quantities, but is not derived from a well defined utility function nor is it guaranteed to be consistent with utility maximization.

used to transform the quantities of products purchased, given by \mathbf{q} into the characteristics that they contain.⁵⁹ The household maximizes utility by choosing \mathbf{q} , subject to a budget constraint:

$$\begin{aligned} & \max_{\mathbf{q}} U(q_0, \mathbf{x}, \mathbf{q}; D_i) \\ \text{s.t.} \quad & \sum_{j=0}^J q_j p_j \leq y_i ; \quad \mathbf{x} = \mathbf{A}'\mathbf{q}; \quad q_j \geq 0, j = 0, 1, \dots, J \end{aligned}$$

where p_j is the price of one unit of product j , and y_i is the household's income.

Following standard arguments (and dropping the i subscripts) this can be written as

$$\begin{aligned} & \max_{\mathbf{q}} U\left(\frac{y - \mathbf{p}'\mathbf{q}}{p_0}, \mathbf{A}'\mathbf{q}, \mathbf{q}\right) \\ \text{s.t.} \quad & q_j \geq 0. \end{aligned}$$

Assuming that quantities $\{q_j\}_{j=0}^J$ are continuous, the first-order conditions are

$$\sum_{k=1}^K a_{jk} \frac{\partial U}{\partial x_k} - \frac{\partial U}{\partial q_0} \frac{p_j}{p_0} + \frac{\partial U}{\partial q_j} = 0 \quad \text{if } q_j > 0.$$

This model nests various standard models. First, suppose the utility function is $U(q_0, \mathbf{x})$, which is the case in discrete choice models or in hedonic models (Court, 1939; Griliches, 1961; Rosen, 1974; Epple, 1987). Because the transformation from products to characteristics is linear and in this case $\partial U / \partial q_j = 0$, at most K of the J products would be purchased. If we restrict $q_j \in \{0, 1\}$ and $\sum_{j=1}^J q_j \leq 1$, the model collapses to the standard discrete choice model. In general, the prediction that at most K products are purchased is a problem since we would like to consider cases where the number of products chosen is (much) greater than the number of observed characteristics.

Alternatively, if the utility function is $U(q_0, \mathbf{q})$ then we can generate standard demand systems in product space. If we allow for a characteristic that is product-specific then a model in characteristics space is equivalent to a model in product space, as long as the characteristics do not vary over time or markets. Note, that we need more than just different values on a small number of unobserved characteristics, but a totally different characteristic that can *only* be obtained from each product. To better understand the role of the characteristics in this model we rewrite the first-order conditions for j such that

⁵⁹This linear transformation is quite natural for characteristics like calories, sugar and protein, where it seems natural to aggregate across products. The linear transformation is less intuitive when it comes to characteristics like ξ in (3.3) that represent demand shocks more generally.

$q_j > 0$ as

$$\frac{\partial U / \partial q_j}{\partial U / \partial q_0} = \frac{p_j}{p_0} - \sum_{k=1}^K a_{jk} \frac{\partial U / \partial x_k}{\partial U / \partial q_0}.$$

Consider the case where characteristics do not enter the utility, i.e., $\partial U / \partial x_k = 0$. The first-order conditions, in this case $\frac{\partial U / \partial q_j}{\partial U / \partial q_0} = \frac{p_j}{p_0}$, which implicitly defines the demand correspondence. A similar idea applies in the above model. Demand depends on the *hedonic prices* of each good instead of prices. The hedonic prices, $\frac{p_j}{p_0} - \sum_{k=1}^K a_{jk} \frac{\partial U / \partial x_k}{\partial U / \partial q_0}$, depend on the marginal utility of the consumer from the characteristics. In other words, if two products have the same price but one has more of a characteristic, with a positive marginal utility, then the effective price to the consumer will be lower for the product with the higher value of the characteristic.

To estimate the model Dubois et al. (2014) focus on a particular functional form for utility and use rich household level data from France, the United Kingdom and the United States. They use the estimated parameters to decompose the cross-country differences into variation coming from differences in preferences and variation coming from the differences in the economic environment.⁶⁰

6.2 Dynamic Demand

Up to this point the models we discussed were static. In this section we explore the implications of dynamic demand. Dynamics in demand can arise for a variety of reasons, including switching costs, learning, storable and durable products, non-separable (over time) utility, pricing (e.g., monthly usage caps) that creates dynamic linkage, and other reasons. Here we will mostly focus on storable and durable products. We will discuss patterns in the data that suggest that dynamic demand is relevant, and discuss the implications of ignoring dynamics. We will then review some of the main modeling and estimation challenges and solutions that have been offered in the literature.

6.2.1 Storable Products

Many of the products that have been studied using the methods presented earlier in this section are storable: in the sense that consumers can buy them in one period and consume in another. Furthermore, a typical pricing pattern in these markets involves short lived price reductions ("sales"), with a return to the regular price. This pattern of prices

⁶⁰Allcott et al. (2019) use a similar demand system, estimated somewhat differently, to study food deserts in the U.S.

generates an incentive for consumers to purchase the product when the price is low and store it for future consumption. Boizot et al. (2001) and Pesendorfer (2002) were among the first to study the effects of temporary price reductions and storability in the economics literature.

The empirical literature has found evidence consistent with the theoretical idea that consumers have incentives to purchase not for immediate consumption but for inventory. For example, Pesendorfer (2002), and Hendel and Nevo (2006b, 2013) find that aggregate quantity sold depends on duration from previous sale, controlling for the current price. Hendel and Nevo (2006b) show that a household's likelihood of purchasing during a temporary price reduction is correlated with proxies of storage costs. They also show that when a household purchases during a temporary price reduction the duration to the next purchase is longer (consistent with the household buying for storage rather than immediate consumption), and households who purchase more often on sale also purchase less frequently overall (again consistent with the idea that these households store their purchases).

Stockpiling behavior has several implications for demand estimates and how they should be used. If consumers purchase for storage, and the evidence suggest that they do, then there is a difference between the short-run response to a temporary price change, and the long-run response to either a temporary or permanent price change. For most economic applications we care about long-run changes. In many data sets temporary price changes account for most of the observed variation in prices. Short-run responses to temporary price reductions, interpreted through a static model, overestimate the long-term own-price effects. Typically, there is a large response to a temporary price reduction, which in a static model is attributed to an increase in consumption (which in a static model equals purchase), and not to an increase in storage. Similarly a post price reduction dip, also often observed in the data, coincides with an increase in price, and is mis-attributed by a static model as a decline in consumption. At the same time, static estimation underestimates cross-price effects: the temporary price reduction diverts current sales from competing products, but it also diverts future sales (and past sales to the extent that the reduction was at least partially anticipated). A static model misses these additional effects and therefore underestimates the impact on other products.

The basic model of consumer stockpiling is an inventory model. Hendel and Nevo (2006b) propose such a model for a homogeneous good, which abstracts from product differentiation and assumes that purchases are of continuous quantities.

The per period utility consumer i obtains from consuming at time t is

$$u_i(c_t + \nu_t) + \alpha_i q_{0t}, \quad (6.1)$$

where c_t is the quantity consumed in t , ν_t is a shock to utility that changes the consumption needs of the consumer, and q_{0t} is the numeraire good consumed at time t . Facing random prices, p_t , in each period the consumer has to decide how much to buy, denoted by q_t , and how much to consume, denoted by c_t . The consumer's problem can therefore be represented as

$$V(\mathbf{s}_t) = \max_{(c_t(\cdot), q_t(\cdot))} \sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E} [u(c_t + \nu_t) - C(i_t) + \alpha p_t q_t \mid \mathbf{s}_t] \quad (6.2)$$

$$s.t. \quad 0 \leq i_t, \quad 0 \leq c_t \quad 0 \leq q_t \quad i_t = i_{t-1} + q_t - c_t,$$

where α is the marginal utility from income, δ is the discount factor, and $C(i_t)$ is the cost of storing inventory. The information set, or state space, at time t , \mathbf{s}_t , consists of the current inventory, i_t , current prices, and the current shock to utility from consumption, ν_t . Consumers face two sources of uncertainty: utility shocks and future prices. Hendel and Nevo assume that shocks to utility, ν_t , are i.i.d. over time. Prices are assumed to evolve according to a first-order Markov process and take on two states, sale and non-sale.

They show that within this setup the optimal consumer behavior is characterized by a trigger s , and a target inventory S . The target, S , is a decreasing function of current price. On the other hand, the trigger, s , which is the sum of the target and current consumption, depends on prices and the utility shock. They also show that the quantity purchased is a function of lagged inventory, the current prices and the current utility shock.

A key challenge is how to expand this model to differentiated products. In principle, this is simple: we can just make all the quantities in (6.1) vectors. This, of course, reintroduces the dimensionality problem discussed in Section 3, except now the problem is even worse and includes the well-known "curse of dimensionality" in dynamic problems (Rust, 1994). To deal with these issues, Hendel and Nevo (2006a) propose the following model.⁶¹ Now consumer i can purchase one of $J + 1$ brands (including a no purchase decision), which come in different sizes, indexed by $q \in \{1, 2, \dots, Q\}$. Let d_{jq_t} equal to 1 if the consumer purchases brand j of size q at time t , and 0 otherwise. Consumers make a

⁶¹See Aguirregabiria and Nevo (2013) for a more detailed discussion.

discrete choice, so in this model stockpiling is achieved by buying larger sizes rather than by buying a larger quantity as in the homogeneous good case.

The consumer also has to decide how much to consume each period.⁶² To reduce the state space Hendel and Nevo assume that the per period utility consumer i obtains from consuming in t is the same as in equation (6.1). This assumption implies that there is no differentiation in consumption: consumers care only about the quantity they consume but not the brand.⁶³ They further assume that inventory cost depends only on total inventory, hence $C(i_t)$ for inventory i_t . They introduce differentiation by assuming an instantaneous utility associated with preference for the purchased brand. At period $t = 1$, the purchase and consumption decisions, $\{c, j, q\}$, are made to maximize

$$\sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E} [u_i(c_t + \nu_t) - C_i(i_t) + x_{jqt}\beta_i + \alpha_i p_{jqt} + \xi_{jqt} + \varepsilon_{ijqt} \mid \mathbf{s}_1] \quad (6.3)$$

$$s.t. \quad 0 \leq i_t, \quad 0 \leq c_t \quad \sum_{j,q} d_{jqt} = 1, \quad i_{t+1} = i_t + \sum_x d_{jqt} q_t - c_{jt} \quad j = 1, \dots, J$$

where \mathbf{s}_t is the information set at time t , p_{jqt} is the price of purchasing quantity q of brand j , ξ_{jqt} is an unobserved (to the researcher) brand-specific quality, x_{jqt} are observed product characteristics and ε_{ijqt} is a random shock. They allow ξ_{jqt} to vary by brand in order to capture differentiation across products, and across sizes.

The expectation $\mathbb{E}(\cdot)$ is taken with respect to the uncertainty regarding future shocks in ν_t and ε_t , and future prices (and other time-varying characteristics). They assume that ε_{ijqt} is i.i.d. type-1 extreme value, and as before that ν_t is i.i.d. over time and across consumers. Prices (and observed characteristics) evolve according to a first-order Markov process.

Let $V_i(\mathbf{s}_t)$ be the value function of consumer i . Given the above assumptions \mathbf{s}_t consists of inventory, i_t , a vector of current prices (and observed characteristics), which we will denote (slightly abusing notation) by \mathbf{x}_t , the scalar shock ν_t and the vector of extreme value shocks ε_{it} . As usual in a dynamic programming problem, this value function can

⁶²Alternatively, one can assume that consumption is constant over time, but varying across households, and not a decision variable. A slightly more general model, than constant consumption, allows for random shocks, that determine consumption. Both these models are nested within the above model and in principle can be tested. The results in Hendel and Nevo (2006a) suggest that consumption is mostly constant, but when inventory runs low consumers reduce consumption. This behavior is required to explain long periods of no purchase followed by periods of frequent purchases observed in the data. Indeed, it is this variation in inter-purchase time that identifies the utility from consumption.

⁶³For the product they study, laundry detergents, this assumption makes sense. This of course raises the question of why products are differentiated. Hendel and Nevo propose an interpretation that allows differentiation in the linear part of the utility.

be obtained as the unique solution of a Bellman equation:

$$V_i(\mathbf{s}_t) = \max_{\{c,j,q\}} \{u_i(c_t + \nu_t) - C_i(i_t) + x_{jqt}\beta_i + \alpha_i p_{jqt} + \xi_{jqt} + \varepsilon_{ijqt} + \delta \int V_i(\mathbf{s}_{t+1}) dF_s(\mathbf{s}_{t+1} | \mathbf{s}_t, c, j, q)\}, \quad (6.4)$$

where F_s represents the transition probability of the vector of state variables. Given that the state variables $(\nu_t, \varepsilon_{it})$ are independently distributed over time, it is convenient to reduce the dimensionality of this dynamic programming problem by using a value function that is integrated over these i.i.d. random variables. The integrated value function, sometimes also called the ex-ante value function, is defined as $EV_i(i_t, \mathbf{x}_t) \equiv \int V_i(\mathbf{s}_t) dF_\varepsilon(\varepsilon_t) dF_\nu(\nu_t)$, where F_ε and F_ν represent the CDFs of ε_t and ν_t , respectively. The value function EV_i is the unique solution of the integrated Bellman equation. Given the distributional assumptions on the shocks ε_t and ν_t , the integrated Bellman equation is:

$$EV_i(i_t, \mathbf{x}_t) = \max_{c,q} \int \ln \left(\sum_j \exp \left\{ \begin{array}{l} u_i(c_t + \nu_t) - C_i(i_t) + x_{jqt}\beta_i + \alpha_i p_{jqt} + \xi_{jqt} \\ + \delta \mathbb{E}[EV_i(i_{t+1}, \mathbf{x}_{t+1}) | i_t, \mathbf{x}_t, c, j, q] \end{array} \right\} \right) dF_\nu(\nu_t). \quad (6.5)$$

Despite the significant reduction in size, the state space is still high dimensional. Therefore, to reduce the dimension further, they note that the assumptions imply that the optimal consumption does not depend on which brand is purchased. Formally, let $c^*(\mathbf{s}_t; q, k)$ be the optimal consumption conditional on state \mathbf{s}_t and on purchase of size q . Lemma 1 in the appendix of Hendel and Nevo (2006a) shows that $c^*(\mathbf{s}_t; q, k) = c^*(\mathbf{s}_t; q, j) = c^*(\mathbf{s}_t; q)$. In words, the optimal consumption does not depend on the brand purchased, only on the size.

This result implies that the (integrated) Bellman equation in (6.5) can be written as

$$EV_i(i_t, \mathbf{x}_t) = \max_{c,q} \int \ln \left(\sum_q \exp \{u_i(c_t + \nu_t) - C_i(i_t) + \omega_{iqt} + \delta \mathbb{E}[EV_i(i_{t+1}, \mathbf{x}_{t+1}) | i_t, \mathbf{x}_t, c, q]\} \right) dF_\nu(\nu_t), \quad (6.6)$$

where ω_{iqt} is the inclusive value from all brands of size q , as defined by equation (3.10), i.e., $\omega_{iqt} = \ln \left(\sum_j \exp(x_{jqt}\beta_i - \alpha_i p_{jqt} + \xi_{jqt}) \right)$. In words, the problem can now be seen as

a choice between sizes, each with a utility given by the size-specific inclusive value (and extreme value shock). The dimension of the state space is still large and includes all characteristics and prices, because we need all the prices to compute the evolution of the inclusive value.

To further reduce the state space Hendel and Nevo assume

$$F(\boldsymbol{\omega}_{i,t+1} \mid \mathbf{s}_t) = F(\boldsymbol{\omega}_{i,t+1} \mid \boldsymbol{\omega}_{it}(\mathbf{x}_t)), \quad (6.7)$$

where $\boldsymbol{\omega}_{it}$ is a vector of inclusive values for the different sizes. In words, the vector $\boldsymbol{\omega}_{it}$ contains all the relevant information in \mathbf{s}_t to obtain the probability distribution of $\boldsymbol{\omega}_{i,t+1}$ conditional on \mathbf{s}_t . Instead of all the prices (and characteristics) we only need a single index for each size. Two vectors of prices (and characteristics) that yield the same (vector of) current inclusive values imply the same distribution of future inclusive values. This assumption is violated if individual prices have predictive power above and beyond the predictive power of $\boldsymbol{\omega}_{it}$. Therefore, if the inclusive values can be estimated outside the dynamic demand model, the assumption can be tested and somewhat relaxed by including additional statistics of prices (and characteristics) in the state space. Note, that $\boldsymbol{\omega}_{it}$ is consumer-specific: different consumers value a given set of products differently and therefore this assumption does not further restrict the distribution of heterogeneity.

Given these assumptions Hendel and Nevo (2006a) show that

$$EV_i(i_t, \mathbf{p}_t) = EV_i(i_t, \boldsymbol{\omega}_{it}(\mathbf{p}_t)) \quad (6.8)$$

In words, the expected future value only depends on a lower dimensional statistic of the full state vector.

Hendel and Nevo estimate the model using consumer level data and using a three-step procedure. First they estimate many of the parameters (including various fixed effects) with a static conditional Logit model where they use the probability of choosing a brand *conditional* on the size being purchased (i.e. they consider only options that have the same size as the size purchased). They show that for this conditional probability they do not need to solve the dynamic programming problem.⁶⁴ Next, they use the first-stage

⁶⁴The intuition for the result is similar to the result in Chamberlain (1980) who proposes to estimate a fixed effects Logit model by conditioning such that the fixed effects drop out. The same happens here, but with the expected value function, instead of a fixed effect.

estimates to compute the transition process of the inclusive values. Finally, they estimate a nested fixed point as in Rust (1987) to estimate the remaining parameters.⁶⁵

They find that estimates that do not account for dynamics overestimate own-price elasticities by roughly 30 percent and underestimate cross-price elasticities by as much as a factor of 5. They also find that static estimates overstate the substitution to the outside option by over 200 percent. Together these suggest that static estimates, like the ones discussed above, might underestimate price–cost margins and be downward biased in predicting the effects of mergers (i.e., static estimates will predict effects that are lower compared to the dynamic model). The models has implications for other policy debates as well. For example, Wang (2015) finds that a static model will overestimate the effect of a soda tax, by as much as 60%.

6.2.2 Durable Products

Many of the papers we discussed above involve estimation of demand for durable goods (Bresnahan, 1987; Berry et al., 1995). Static models miss two important dynamic effects. First, whether a product is owned (and which one) is likely to impact purchases. For example, a consumer who more recently purchased a cell phone might be less likely to buy a new phone than a consumer who owns an older model. Second, purchase decisions will depend on expectations about future prices and quality. Expectation about the future are especially important when nominal prices are declining and quality increasing, as is the case in many durable good industries. The decline in quality-adjusted prices creates a trade off for consumers between purchasing today, and getting the benefits of usage earlier, or delaying purchase and paying a lower price (or getting higher quality).

Initially, the literature separated the modeling between two cases: with and without repeat purchase, as far as we can tell mostly because the no repeat purchase was easier to deal with. More recently the literature has focused on the repeat purchase case, which seems to better fit reality, and that is the one we mostly focus on here. With repeat purchases the main issues with the static model are the ones discussed in the previous paragraph. It is difficult to theoretically sign the direction of the bias in a static model, but empirically it seems like static estimates are lower in absolute value.⁶⁶

⁶⁵They need to modify the Rust estimation algorithm to account for the fact that inventory, a state variable, is unobserved.

⁶⁶Without repeat purchase the issues with the static model are a bit different. First, after consumers purchase they leave the market, and if consumers are heterogeneous then the distribution of the remaining consumers changes over time in a way that is not accounted for by the static model. Second, if consumers

Gowrisankaran and Rysman (2012) offer a framework that extends the static BLP model and allows for dynamics. Interestingly, their model in several ways is similar to the inventory model we presented in the previous section, where the role of inventory is equivalent to the role of the quality of the product owned, in the model below. In the durable good model "stockpiling" means buying a higher quality product, i.e., "stockpiling" quality rather than quantity. The real difference between the two models is in the pricing patterns and therefore the trade-off faced by consumers. In storable goods markets, consumers face temporary price reductions that create an incentive to purchase today for future consumption. In durable goods markets the typical pattern is a decreasing quality-adjusted price, which creates an incentive to delay purchase, either by not buying today or by buying a lower quality product, with the intention of replacing it soon, or renting/leasing.⁶⁷

To model these effects, let the (indirect) utility consumer i gets from product j at time t be given by:

$$u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}, \quad (6.9)$$

where the notation follows the definitions of the static model in Section 3. For what follows it is convenient to define the flow utility as $\gamma_{ijt}^f = x_{jt}\beta_i + \xi_{jt}$. If the consumer does not purchase she gets the utility $u_{i0t} = \gamma_{i0t}^f + \varepsilon_{i0t}$ where

$$\gamma_{i0t}^f = \begin{cases} 0 & \text{if no previous purchase} \\ \gamma_{i\hat{j}t}^f & \text{if last purchase was product } \hat{j} \text{ at time } \hat{t} \end{cases}. \quad (6.10)$$

This definition of the utility from the outside option is the main difference between the static model and the dynamic model. In the static model, the utility from the outside good is constant (and normalized to zero) across consumers and over time. In the dynamic model, once consumers purchase, the utility from the outside option changes. Forward-looking consumers take this into account when making current choices. Assuming that (i) the consumer holds at most a single product at any time and (ii) there is no resale market,

are forward-looking then they realize there is an option value to not purchasing today. This option value is reflected in the value of the outside option, which in the static model is assumed constant. Melnikov (2013) and Conlon (2012) offer models and empirical estimates of the no-repeat purchase model.

⁶⁷In some cases, dynamics can also arise due to temporary price changes in durable goods markets. For example, Busse et al. (2010) study the 2005 Employee Discount Pricing, and show that its main effect was to induce consumers to purchase earlier.

then the Bellman equation of the consumer problem is given by

$$V_i(\boldsymbol{\varepsilon}_{it}, \gamma_{i0t}^f, \mathbf{x}_t) = \max_{j=0, \dots, J} \left\{ u_{ijt} + \delta \mathbb{E}[EV_i(\gamma_{ijt+1}^f, \mathbf{x}_{t+1}) | \mathbf{x}_t] \right\}, \quad (6.11)$$

where $EV_i(\gamma_{ijt}^f, \mathbf{x}_t) = \int V_i(\boldsymbol{\varepsilon}_{it}, \gamma_{ijt}^f, \mathbf{x}_t) dF_\varepsilon(\boldsymbol{\varepsilon}_{it})$, and \mathbf{x}_t represents the set of prices and other product characteristics at period t . The expectation is taken with respect to the uncertainty regarding future products, prices and characteristics.

Note, that there is another similarity with the storable goods model. Here the utility carried forward is γ_{ijt}^f and not $\gamma_{ijt}^f + \varepsilon_{ijt}$. Thus, just like in the inventory model there is a separation between the utility at the time of purchase and the utility at the time of usage.

As is usually the case, it is convenient to work with the integrated value function. Even with this, the state space includes the vector of all characteristics and prices, which is still too large to practically work with. They reduce the state space in a similar way to what we saw in the storable goods model. The state space also includes the quality of the products currently held, which is equivalent to the inventory in the storable good problem. Because they assume that the consumer only holds a single product, this quality is a scalar. In more general models the consumer might purchase or hold multiple products, or multiple units of the same product, and the dimension of quality would be higher. To reduce the dimension of the state space, they rely, similar to the storable goods model, on the inclusive value. However, it will be defined slightly differently here. Specifically, define the dynamic inclusive value from the J inside alternatives as:

$$\omega_{it}(\mathbf{x}_t) = \ln \left(\sum_{j=1}^J \exp(\gamma_{ijt}^f + \alpha_i p_{jt} + \delta \mathbb{E}[EV_i(\gamma_{ijt+1}^f, \mathbf{x}_{t+1}) | \mathbf{x}_t]) \right). \quad (6.12)$$

Note, that this definition is different in an important way from the definition given in equation (3.10). The above definition provides the expected value, including the future value, from the J options, while the definition in equation (3.10) provides the expected flow utility, not accounting for the future value. The difference is not just semantic. The static definition basically provides a (utility-consistent welfare) statistic that is a summary of prices and characteristics of available products. The dynamic definition also includes (endogenous) future behavior of the agent. Once we impose a particular stochastic structure on the evolution of ω_{it} a natural question is whether the imposed structure is consistent with the consumer optimization problem. Gowrisankaran and Rysman (2012) offer

some discussion on whether or not this is restrictive, but generally little is known on what behavioral assumptions are consistent with the imposed structure.

Given this definition, Gowrisankaran and Rysman make a similar assumption to equation (6.7) made in the storable goods model

$$F(\omega_{i,t+1} \mid \mathbf{x}_t) = F(\omega_{i,t+1} \mid \omega_{it}(\mathbf{x}_t)). \quad (6.13)$$

As before, the assumption is that the inclusive value is sufficient to compute the transition probabilities, but now it is the dynamic inclusive value, ω_{it} . Furthermore, now there is a single inclusive value rather than a vector of size-specific inclusive values, as was the case for the stockpiling model.

Using this assumption we can now write

$$EV_i(\gamma_{i0t}^f, \mathbf{x}_t) = EV_i(\gamma_{i0t}^f, \omega_{it}) = \ln \left(\exp(\omega_{it}) + \exp \left(\gamma_{i0t}^f + \delta \mathbb{E}[EV_i(\gamma_{i0t+1}^f, \omega_{it+1} \mid \omega_{it})] \right) \right). \quad (6.14)$$

Several studies that have estimated demand for durable products using household level data.⁶⁸ However, Gowrisankaran and Rysman (2012) offer a way to estimate the model using aggregate data, which directly extend the methods of BLP.

If consumer level data are observed then, in principle, identification follows the standard arguments (Rust, 1994; Magnac and Thesmar, 2002).⁶⁹ With aggregate data we do not observe the purchase history of each consumer, which makes identification significantly more difficult. Intuitively, the key to identifying the model and to separating the different alternative models is the ability of the models to explain both the cross-sectional variation, across markets and products, and the time series variation.

We are unaware of a formal identification proof. Standard identification proofs for static models require some form of substitution between products (Berry et al., 2013). In static models the substitution is between products in a given period, but here the requirement is for substitution over time and across products. This need not be satisfied. For example, if the price of a high quality product falls at time t it could actually increase the demand for a low quality product at $t - 1$, because some consumers might buy it for one period.

The estimation, using aggregate data, follows closely the method proposed by Berry et al. (1995), but nests a solution of the dynamic programming problem inside the inner

⁶⁸For example, Erdem et al. (2005), or Prince (2008).

⁶⁹The standard arguments need to be adjusted for the existence of ξ_{jt} , but with enough observations these could be controlled for and then we are back in the standard case.

loop. The idea is to follow the algorithm detailed in Section 4.3.4, but in Step 1 in order to compute the market shares we need to solve the dynamic problem for each of the simulated individuals. This is done by computing the inclusive value (using equation (6.12) and an initial guess for EV_i) for each simulated individual i . This in turn is used to compute $F(\omega_{it+1} | \omega_{it}(\mathbf{x}_t))$, which is used to update EV_i . The process is continued until it converges. If we think of the BLP algorithm as a nested fixed point, then here we have another layer of nesting in order to compute the market shares.

7 Concluding Comments

In this chapter we review the modern IO approach to modeling demand and supply in differentiated products industries. In many cases, we only scratch the surface of many of the topics we discussed and there are many topics that are left uncovered. For example, many of the applications of the models we discussed are covered in other chapters of this Handbook. The success of the methods we discuss here is reflected in their application to areas such as health, finance, taxation, housing and school choice, development, environmental policy and political economy, that historically were very different than IO. This is a positive trend that we hope will continue into the future as IO economists continue developing more flexible models and improved computational methods.

References

- Akerberg, D., L. Benkard, S. Berry, and A. Pakes (2007). *Econometric Tools for Analyzing Market Outcomes*, Volume 6A, pp. 4171–4276. Amsterdam: North-Holland. Draft Date: June, 2006.
- Akerberg, D. A. and M. Rysman (2005). Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *The RAND Journal of Economics* 36(4), 771–788.
- Aguirregabiria, V. and A. Nevo (2013). Recent developments in empirical io: Dynamic demand and dynamic games. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics: Tenth World Congress*, Volume 3 of *Econometric Society Monographs*, pp. 53–122. Cambridge University Press.
- Allcott, H., R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell (2019). Food Deserts and the Causes of Nutritional Inequality. *The Quarterly Journal of Economics* 134(4), 1793–1844.
- Anderson, S. P., A. de Palma, and J.-F. Thisse (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press.
- Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *International Economic Review* 58(2).
- Athey, S. and G. Imbens (2007). Discrete choice models with multiple unobserved choice characteristics. *International Economic Review* 48(4), 1159–1192.
- Backus, M., C. Conlon, and M. Sinkinson (2021). Common ownership and competition in the ready-to-eat cereal industry. Working Paper 28350, National Bureau of Economic Research.
- Bagwell, K., R. W. Staiger, and A. Yurukoglu (2020). Quantitative analysis of multi-party tariff negotiations estimation of a model of entry in the airline industry. *Working paper*.
- Bain, J. S. (1951). Relation of profit rate to industry concentration: American manufacturing, 1936–1940. *The Quarterly Journal of Economics* 65(3), 293–324.

- Barten, A. (1966). *Theorie en Empirie van een Volledig Stelsel van Vraagvergelijkingen*, Doctoral dissertation. Rotterdam: University of Rotterdam.
- Bayot, D., K. Hatzitaskos, B. Howells, and A. Nevo (2018). The aetna-humana proposed merger. In *The Antitrust Revolution, Volume 7*. Oxford University Press.
- Berry, S. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.
- Berry, S., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica* 81(5), 2087–2111.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63(4), 841–890.
- Berry, S., J. Levinsohn, and A. Pakes (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review* 89(3), 400–430.
- Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new vehicle market. *Journal of Political Economy* 112, 68–104.
- Berry, S., O. Linton, and A. Pakes (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *Review of Economic Studies* 71(3), 613–654.
- Berry, S. and A. Pakes (1993). Some applications and limitations of recent advances in empirical industrial organization: Merger analysis. *American Economic Review* 83(2), 247–252.
- Berry, S. and A. Pakes (2007). The pure characteristics demand model. *International Economic Review* 48(4).
- Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica* 82(5), 1749–1797.
- Berry, S. T. and P. A. Haile (2021). Foundations of demand estimation. Volume 4 of *Handbook of Industrial Organization*. Elsevier.
- Bhattacharya, D. (2018). Empirical welfare analysis for discrete choice: Some general results. *Quantitative Economics* 9(2), 571–615.

- Binmore, K., A. Rubinstein, and A. Wolinsky (1986). The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 176–188.
- Black, R., G. S. Crawford, S. Lu, and H. White (2004). A virtual stakes approach to measuring competition in product markets. *Technical Paper*.
- Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.
- Boizot, C., J.-M. Robin, and M. Visser (2001). The demand for food products: An analysis of interpurchase times and purchased quantities. *The Economic Journal* 111(470), 391–419.
- Bonnet, C. and P. Dubois (2010). Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance. *The RAND Journal of Economics* 41(1), 139–164.
- Boyd, J. H. and R. E. Mellman (1980). The effect of fuel economy standards on the u.s. automotive market: An hedonic demand analysis. *Transportation Research* 14a, 367–368.
- Bresnahan, T. (1981). Departures from marginal-cost pricing in the american automobile industry: Estimates for 1977-1978. *Journal of Econometrics* 17(2), 201–227.
- Bresnahan, T. F. (1982). The oligopoly solution concept is identified. *Economics Letters* 10(1), 87 – 92.
- Bresnahan, T. F. (1987). Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics* 35(4), 457–482.
- Bresnahan, T. F. (1989). Chapter 17 empirical studies of industries with market power. Volume 2 of *Handbook of Industrial Organization*, pp. 1011 – 1057. Elsevier.
- Busse, M. R., D. I. Simester, and F. Zettelmeyer (2010). “The Best Price You’ll Ever Get”: The 2005 employee discount pricing promotions in the u.s. automobile industry. *Marketing Science* 29(2), 268–290.
- Capps, C., D. Dranove, and M. Satterthwaite (2003). Competition and market power in option demand markets. *RAND Journal of Economics* 34(4), 737–763.

- Cardell, N. and F. Dunbar (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research 14a*, 423–434.
- Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory 13*(2), 185–213.
- Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies 47*(1), 225–238.
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics 34*(3), 305–334.
- Christensen, L., D. Jorgenson, and L. Lau (1975). Transcendental logarithmic utility functions. *American Economic Review 65*, 367–83.
- Ciliberto, F. and J. W. Williams (2014). Does multimarket contact facilitate tacit collusion? inference on conduct parameters in the airline industry. *The RAND Journal of Economics 45*(4), 764–791.
- Collard-Wexler, A., G. Gowrisankaran, and R. S. Lee (2019). *Journal of Political Economy 127*(1).
- Compiani, G. (2019). Market counterfactuals and the specification of multi-product demand: a nonparametric approach. Technical report, Working paper. 3, 43.
- Conlon, C. (2012). A dynamic model of prices and margins in the lcd tv industry. *Working paper*.
- Conlon, C. and J. Gortmaker (2020). Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics 51*(4), 1108–1161.
- Corts, K. S. (1999). Conduct parameters and the measurement of market power. *Journal of Econometrics 88*(2), 227–250.
- Court, A. T. (1939). Hedonic price indexes with automotive examples. *The Dynamics of Automotive Demand*, edited by Charles F. Roos, pp. 99–117. New York: General Motors.

- Crawford, G. S. and A. Yurukoglu (2012). The welfare effects of bundling in multichannel television Markets. *American Economic Review* 102(2), 643–685.
- Davidson, R. and J. G. MacKinnon (1989). Testing for consistency using artificial regressions. *Econometric theory*, 363–384.
- Deaton, A. (1986). Demand analysis. Volume 3 of *Handbook of Econometrics*, pp. 1767–1839. Elsevier.
- Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *American Economic Review* 70, 312–26.
- Debreu, G. (1960). Review of R.D. Luce, individual choice behavior: A theoretical analysis. *American Economic Review* 50, 186–188.
- Dixit, A. and J. Stiglitz (1977). Monopolistic competition and optimum product diversity. *American Economic Review* 67, 297–308.
- Donald, S. G., G. W. Imbens, and W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* 117(1), 55–93.
- Duarte, M., L. Magnolfi, M. Sølvssten, and C. Sullivan (2021). Testing firm conduct. Working paper.
- Dubé, J.-P. (2004). Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science* 23, 66–81.
- Dubé, J.-P. (2019). Microeconomic models of consumer demand. In J.-P. Dubé and P. Rossi (Eds.), *Handbook of the Economics of Marketing, Volume 1*. Elsevier.
- Dubé, J.-P., J. T. Fox, and C.-L. Su (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica* 80(5), 2231–2267.
- Dubé, J.-P., A. Hortaçsu, and J. Joo (2021). Random-coefficients logit demand estimation with zero-valued market shares. *Marketing Science*.
- Dubin, J. A. and D. McFadden (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* 52(2), 345–62.

- Dubois, P., R. Griffith, and A. Nevo (2014). Do prices and attributes explain international differences in food purchases? *American Economic Review* 104(3), 832–67.
- Dubois, P., R. Griffith, and M. O’Connell (2020). How well targeted are soda taxes? *American Economic Review* 110(11), 3661–3704.
- Epple, D. (1987). Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy* 95(1), 59–80.
- Erdem, T., M. P. Keane, T. S. Öncü, and J. Strebel (2005). Learning about computers: An analysis of information search and technology choice. *Quantitative Marketing and Economics* 3(3), 207–247.
- Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review* 103(5), 1598–1628.
- Fosgerau, M., J. Monardo, and A. De Palma (2020). The inverse product differentiation logit model.
- Fox, J. T. and A. Gandhi (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *The RAND Journal of Economics* 47(1), 118–139.
- Fox, J. T., K. il Kim, and C. Yang (2016). A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics* 195(2), 236–254.
- Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics* 2(3), 381–418.
- Freyberger, J. (2015). Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics* 181(1), 162–181.
- Gandhi, A. and J.-F. Houde (2020). Measuring substitution patterns in differentiated products industries. *Working Paper*.
- Gandhi, A., Z. Lu, and X. Shi (2021). Estimating demand for differentiated products with zeroes in market share data. *working paper*.

- Gandhi, A., A. Nevo, and J. Tao (2021). Flexible estimation of differentiated product demand models using aggregate data. *Working paper*.
- Gasmi, F., J. Laffont, and Q. Vuong (1992). Econometric analysis of collusive behavior in a soft-drink market. *Journal of Economics & Management Strategy* 1(2), 277–311.
- Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review* 97(3), 713–744.
- Ghili, S. (Forthcoming). Network formation and bargaining in vertical markets: The case of narrow networks in health insurance. *Marketing Science*.
- Gorman, W. (1956). A possible procedure for analysing quality differentials in the egg market. *Reprinted in The Review of Economic Studies* (1980).
- Gorman, W. (1980). A possible procedure for analysing quality differentials in the egg market. *Review of Economic Studies* 47(5), 843–856.
- Gowrisankaran, G., A. Nevo, and R. Town (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review*, 172–203.
- Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy* 120(6), 1173–1219.
- Grennan, M. (2013). Price discrimination and bargaining: Empirical evidence from medical devices. *American Economic Review* 103(1), 145–77.
- Grieco, P. L. E., C. Murry, J. Pinkse, and S. Sagl (2021). Efficient estimation of random coefficients demand models using product and consumer datasets. *Working Paper*.
- Griffith, R. and A. Nevo (2019). Marketing and public policy. In J.-P. Dubé and P. Rossi (Eds.), *Handbook of the Economics of Marketing, Volume 1*. Elsevier.
- Griliches, Z. (1961). Hedonic price indexes for automobiles: An econometric analysis of quality change. In *The Price Statistics of the Federal Government*, pp. 173–196. Cambridge, MA: National Bureau of Economic Research.
- Hanemann, W. M. (1984). Discrete/continuous models of consumer demand. *Econometrica* 52(3), 541–561.

- Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In T. Bresnahan and R. Gordon (Eds.), *The Economics of New Goods*, Volume 58 of *Studies in Income and Wealth Vol. 58*. Chicago: National Bureau of Economic Research.
- Hausman, J., G. Leonard, and J. D. Zona (1994). Competitive Analysis with Differentiated Products. *Annales D'Economie et de Statistique* 34, 159–180.
- Hendel, I. (1999). Estimating multiple-discrete choice models: An application to computerization returns. *The Review of Economic Studies* 66(2), 423–446.
- Hendel, I. and A. Nevo (2006a). Measuring the implications of sales and consumer inventory behavior. *Econometrica* 74, 1637–1674.
- Hendel, I. and A. Nevo (2006b). Sales and consumer inventory. *The RAND Journal of Economics* 37(3), 543–561.
- Hendel, I. and A. Nevo (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review* 103(7), 2722–51.
- Ho, K. and R. S. Lee (2019). Equilibrium insurer-provider networks: Bargaining and exclusion in health care markets. *American Economic Review* 109(2), 473–522.
- Hong, H., H. Li, and J. Li (2021). BLP estimation using laplace transformation and overlapping simulation draws. *Journal of Econometrics* 222(1), 56–72.
- Honka, E., A. Hortaçsu, and M. Wildenbeest (2019). Empirical search and consideration sets. In J.-P. Dubé and P. E. Rossi (Eds.), *Handbook of the Economics of Marketing, Volume 1*, Volume 1 of *Handbook of the Economics of Marketing*, pp. 193–257. North-Holland.
- Horn, H. and A. Wolinsky (1988a). Bilateral monopolies and incentives for merger. *RAND Journal of Economics* 19(3), 408–419.
- Horn, H. and A. Wolinsky (1988b). Worker substitutability and patterns of unionisation. *The Economic Journal* 98(391), 484–497.
- Knittel, C. R. and K. Metaxoglou (2014). Estimation of random-coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics* 96, 34–59.
- Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy* 74(2).

- Lau, L. J. (1982). On identifying the degree of competitiveness from industry price and output data. *Economics Letters* 10(1-2), 93–99.
- Lee, J. and K. Seo (2015). A computationally fast estimator for random coefficients logit demand models using aggregate data. *The RAND Journal of Economics* 46(1), 86–102.
- Lee, R., A. Yurukoglu, and M. Whinston (2021). Volume 4 of *Handbook of Industrial Organization*. Elsevier.
- Lee, R. S. and K. Fong (2013). Markov-perfect network formation an applied framework for bilateral oligopoly and bargaining in buyer-seller networks. *Working paper*.
- Liebman, E. (2018). Bargaining in markets with exclusion: An analysis of health insurance networks. *Working Paper*.
- Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete decision processes. *Econometrica* 70(2), 801–816.
- Malone, J., A. Nevo, and J. Williams (2020). The tragedy of the last mile: Congestion externalities in broadband networks. *Working Paper*.
- McFadden, D. (1978). Modeling the choice of residential location. in A. Karlqvist, et al., eds., *Spatial Interaction Theory and Planning Models*.
- McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications* 198272.
- McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15(5), 447–470.
- Melnikov, O. (2013). Demand for differentiated durable products: The case of the u.s. computer printer market. *Economic Inquiry* 51(2), 1277–1298.
- Miller, N. H. and M. C. Weinberg (2017). Understanding the price effects of the millercoors joint venture. *Econometrica* 85(6), 1763–1791.
- Nevo, A. (1998). Identification of the oligopoly solution concept in a differentiated-products industry. *Economics Letters* 59(3), 391 – 395.
- Nevo, A. (2000a). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *RAND Journal of Economics* 31, 395–421.

- Nevo, A. (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, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2), 307–342.
- Nevo, A. (2003). New products, quality changes and welfare measures computed from estimated demand systems. *The Review of Economics and Statistics* 85(2), 266–275.
- Nevo, A. (2011). Empirical models of consumer behavior. *Annual Reviews*.
- Nevo, A., D. L. Rubinfeld, and M. McCabe (2005). Academic journal pricing and the demand of libraries. *American Economic Review* 95(2), 447–452.
- Nevo, A., J. Turner, and J. Williams (2016). Usage-based pricing and demand for residential broadband. *Econometrica* 84(2), 411–443.
- Pakes, A. (2017). Empirical tools and competition analysis: Past progress and current problems. *International Journal of Industrial Organization* 53, 241–266.
- Pakes, A., S. Berry, and J. A. Levinsohn (1993). Applications and limitations of some recent advances in empirical industrial organization: Price indexes and the analysis of environmental change. *American Economic Review* 83(2), 240–246.
- Pesendorfer, M. (2002). Retail sales: A study of pricing behavior in supermarkets. *The Journal of Business* 75(1), 33–66.
- Peters, C. (2006). Evaluating the Performance of Merger Simulations: Evidence from the U.S. Airline Industry. *Journal of Law and Economics* 47(3), 627–649.
- Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 705–29.
- Pinkse, J. and M. Slade (2004). Mergers, brand competition, and the price of a pint. *European Economic Review* 48(3), 617–643.
- Pinkse, J., M. E. Slade, and C. Brett (2002). Spatial price competition: A semiparametric approach. *Econometrica* 70(3), 1111–1153.

- Prince, J. T. (2008). Repeat purchase amid rapid quality improvement: Structural estimation of demand for personal computers. *Journal of Economics & Management Strategy* 17(1), 1–33.
- Reiss, P. C. and F. A. Wolak (2007). *Structural Econometric Modeling: Rationales and Examples from Industrial Organization*, Volume 6A. Amsterdam: North-Holland.
- Reynaert, M. and F. Verboven (2014). Improving the performance of random coefficients demand models: the role of optimal instruments. *Journal of Econometrics* 179(1), 83–98.
- Rivers, D. and Q. Vuong (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal* 5(1), 1–39.
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 34–55.
- Rosse, J. N. (1970). Estimating cost function parameters without using cost data: Illustrated methodology. *Econometrica*, 256–275.
- Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica* (50), 97–109.
- Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica* 55(5), 999–1033.
- Rust, J. (1994). Chapter 51 structural estimation of markov decision processes. Volume 4 of *Handbook of Econometrics*, pp. 3081 – 3143. Elsevier.
- Salanié, B. and F. A. Wolak (2019). Fast, “robust”, and approximately correct: estimating mixed demand systems. Technical report, National Bureau of Economic Research.
- Scherer, F. M. (1982). The breakfast cereal industry. In W. Adams (Ed.), *The Structure of American Industry*. New York: Macmillan.
- Schultz, H. (1938). *The Theory and Measurement of Demand*. Chicago: The University of Chicago Press.
- Shaked, A. and J. Sutton (1983). Natural oligopolies. *Econometrica* 51(5), 1469–1483.
- Small, K. A. and H. S. Rosen (1981). Applied welfare analysis with discrete choice models. *Econometrica* 49, 105–30.

- Smith, H. (2004). Supermarket choice and supermarket competition in market equilibrium. *Review of Economic Studies* 71(1), 235–263.
- Smith, R. J. (1992). Non-nested tests for competing models estimated by generalized method of moments. *Econometrica* 60(4), 973–980.
- Spence, M. (1976). Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies* 43, 217–235.
- Stigler, G. (1954). The early studies of empirical studies of consumer behavior. *The Journal of Political Economy* 62, 95–113.
- Stone, J. (1954). *The Measurement of Consumer Expenditure and Behavior in the United Kingdom, 1920-1938, Vol 1*. Cambridge University Press.
- Sweeting, A. (2013). Dynamic product positioning in differentiated product industries: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica* 81(5), 1763–1803.
- Theil, H. (1965). The information approach to demand analysis. *Econometrica* 6, 375–80.
- Thomassen, Ø., H. Smith, S. Seiler, and P. Schiraldi (2017). Multi-category competition and market power: A model of supermarket pricing. *American Economic Review* 107(8), 2308–51.
- Town, R. and S. Liu (2003). The welfare impact of medicare hmos. *The RAND Journal of Economics* 34(4), 719–736.
- Town, R. and G. Vistnes (2001). Hospital competition in HMO networks. *Journal of Health Economics* 20(5), 733–752.
- Trajtenberg, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy* 97, 444–79.
- Villas-Boas, S. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *Review of Economic Studies* 74, 625–652.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–33.

- Waldfogel, J. (2003). Preference externalities: An empirical study of who benefits whom in differentiated-product markets. *RAND Journal of Economics* 34(3), 557–68.
- Wang, E. Y. (2015). The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. *The RAND Journal of Economics* 46(2), 409–441.
- Werden, G. J. and L. M. Froeb (1994). The effects of mergers in differentiated products industries: Logit demand and merger policy. *Journal of Law, Economics, & Organization* 10(2), 407–426.
- Wollmann, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review* 108(6), 1364–1406.