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.

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

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 Ackerberg 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.

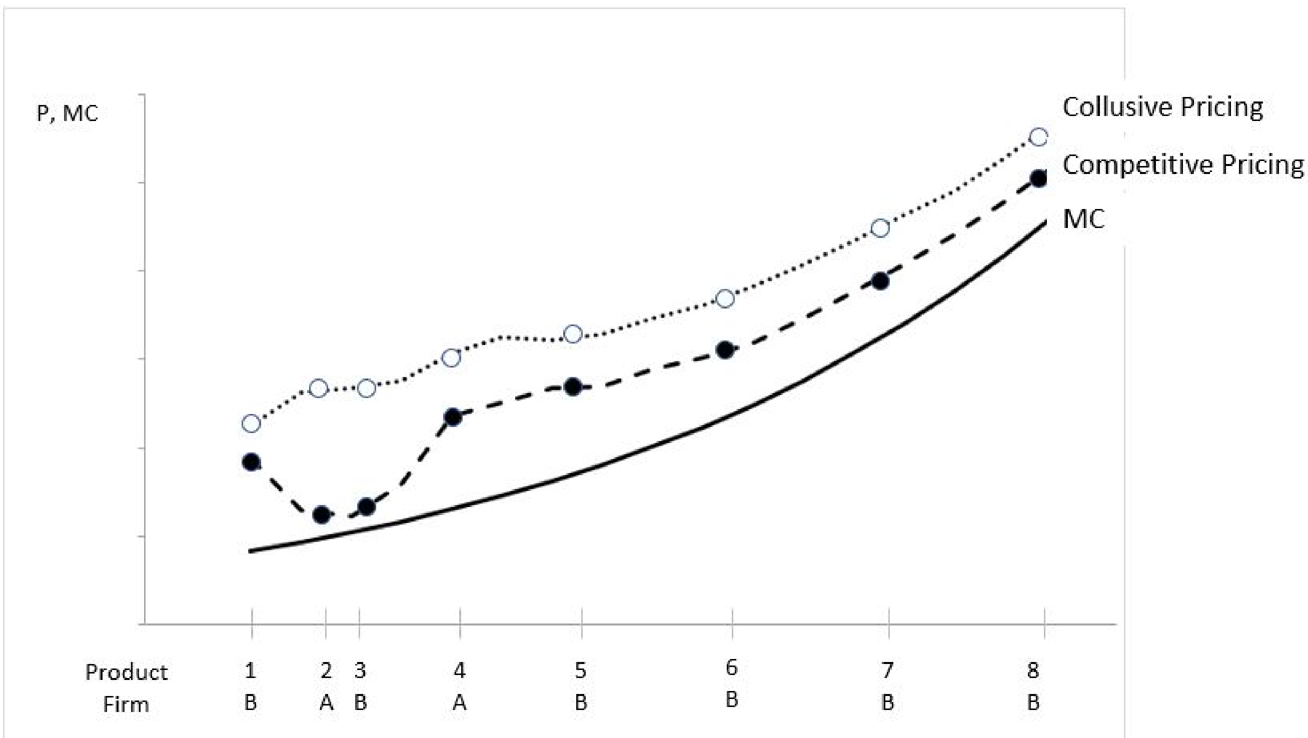
To infer the model of competition, he uses variation in demand conditions across different cars. The basic idea can be seen by examining Figure 2.1, which is a modified rendering of Figure 2 in Bresnahan (1987). Each point in the figure represents a product produced and priced by one of two firms, A 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 as 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

[1] See 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 rending 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 $f = 1, ..., F$ denote firms and $j = 1, ..., J$ 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 $\boldsymbol{p}$ 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

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

where $q_j(\boldsymbol{p})$ 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

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

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

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

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

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

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,

[2]For 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

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

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

$$\nu_i \cdot quality_j + y_i - p_j,$$

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

[3] The 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 $j = 1, \ldots, J+1$ 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

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

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, $x_j^{(k)}$, $k = 1, \ldots, K$, 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 $(\beta_0, \ldots, \beta_K)$ 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

[4]For 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

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

where $p^*(\boldsymbol{x}; H, \theta)$ and $q^*(\boldsymbol{x}; H, \theta)$ are the equilibrium prices and quantities predicted by the model, $\boldsymbol{x}$ 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.

[5]Specifically, 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.

[6]The 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

$$q_{jt} = Q_j(\boldsymbol{p}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t), \qquad j = 1, ...J, \tag{3.1}$$

where q_{jt} is the quantity demanded of product j in market t, $Q_j(\cdot)$ is the demand function for product j, $\boldsymbol{p}_t$ is a $J \times 1$ vector of prices, $\boldsymbol{x}_t$ 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,

$$\boldsymbol{q}_t = A\boldsymbol{p}_t + \boldsymbol{\epsilon}\left(\boldsymbol{\xi}_t\right) \tag{3.2}$$

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

where $\boldsymbol{q}_t$ is a $J \times 1$ vector of quantities, A is $J \times J$ matrix of parameters and $\boldsymbol{\epsilon}(\boldsymbol{\xi})$ 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, $\boldsymbol{x}_t$, 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 microfounding 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 Stiglitiz, 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

[8]This 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, $i = 1, \ldots, I_t$ in market t, $t = 1, \ldots, T$, from product j, $j = 1, \ldots, J$[9] is given by

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

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, $\boldsymbol{x}_t$, $\boldsymbol{p}_t$ 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

[9] To 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

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

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), $D_{it} = (D_{i1t}, ..., D_{iLt})^\top$, 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

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

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

$$\delta_{jt} = x_{jt}\beta_0 + \alpha_0 p_{jt} + \xi_{jt} \tag{3.6}$$

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 $(\alpha_v, \beta_v^{(1)}, ..., \beta_v^{(K)})$, and $\nu_{it} = (\nu_{it}^{(0)}, ..., \nu_{it}^{(K)})^\top$. 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

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

[10] In 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 $\boldsymbol{\delta}_t = (\delta_{1t}, \dots, \delta_{Jt})$.

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

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

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

[11] In 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, $g = 1, ..., G$. Finally, let $\varepsilon_{ijt} = \lambda\varepsilon_{ig(j)t} + \varepsilon^1_{ijt}$, where ε^1_{ijt} 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

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

[12] In 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.

[13] Note, 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 $(\varepsilon_{i0t}, ...\varepsilon_{iJt})$. The expectation needs to account for a selection problem: the choice maximizes the utility given in equation (3.3) after observing $(\varepsilon_{i0t}, ...\varepsilon_{iJt})$. 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

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

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

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

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

[15] In 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

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

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-

[16]Nevo (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