EMPIRICAL MODELS OF CONSUMER BEHAVIOR

Aviv Nevo

Empirical Models of Consumer Behavior
Aviv Nevo
NBER Working Paper No. 16511
October 2010
JEL No. C01,L0

ABSTRACT

Models of consumer behavior play a key role in modern empirical Industrial Organization. In this paper, I survey some of the models used in this literature. In particular, I discuss two commonly used demand systems: multi-stage budgeting approaches and discrete choice models. I motivate their use and highlight some key modeling assumptions. I next briefly discuss key issues of estimation, and conclude by summarizing some extensions.

Aviv Nevo
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, IL 60208-2600
and NBER
nevo@northwestern.edu

Empirical Models of Consumer Behavior

Aviv Nevo[*]

October 10, 2010

Abstract

Models of consumer behavior play a key role in modern empirical Industrial Organization. In this paper, I survey some of the models used in this literature. In particular, I discuss two commonly used demand systems: multi-stage budgeting approaches and discrete choice models. I motivate their use and highlight some key modeling assumptions. I next briefly discuss key issues of estimation, and conclude by summarizing some extensions.

Keywords: Industrial Organization; Demand Estimation; Differentiated Products; Almost Ideal Demand System; Discerte Choice;

1 Introduction

The empirical analysis of consumer behavior has a long and rich history in economics and econometrics. The first statistical estimation of demand dates back at least to Moore (1914).[1] Early work treated estimation as merely a way of summarizing data, and had little connection with economic theory. Since the pioneering work of Stone (1954) econometricians estimating demand systems have struggled with the need for flexible functional forms, which do not impose a prior the data cannot overcome, while keeping a connection to economic theory (either by imposing it, or finding ways to test it). Examples include the Rotterdam model (Theil, 1965; and Barten 1966), the Translog model (Christensen, Jorgenson, and Lau, 1975), and the Almost Ideal Demand System (Deaton and Muellbauer, 1980a). Deaton (1986) offers a comprehensive review of this literature.

---

[1]Moore's work was pre-dated by attempts to summarize relations between quantities and prices, see Schultz (1938) and Stigler (1954) for a survey of the early work and a discussion of Moore's contributions.

A parallel line of research treats goods as bundles of attributes, rather than qualitatively different products (Gorman, 1980, Lancester 1966 and Rosen, 1974). Within this class of characteristics based models especially prevalent is the study of discrete choice (McFadden, 1974), which like the work on demand model, also emphasizes the direct and close connection between economic theory, econometrics and empirical work. See McFadden (1981, 1984) and Train (2003) for surveys of this line of research.

Since the mid 1980's, however, many researchers in some fields of applied micro econometrics have lost interest in estimating consumer behavior. Instead, the focus, in some empirical fields, shifted to estimation of so called casual, or treatment effects, models using natural and quasi experiments. This shift was not uniform across and within all fields of micro economics. Industrial Organization (IO) is one of the fields where empirical analysis of consumer behavior gained prominence during this period. Estimation of demand for differentiated products plays a key role in modern empirical IO. Indeed, several of the recent developments in the study of consumer behavior have been within the field of IO, which might seem out of place since IO is historically mainly associated with study of competition and the supply side.

IO economists' are interested in estimating consumer behavior for several reasons. Two leading examples are to infer firm conduct and to measure (changes in) consumer welfare. An important part of IO involves trying to understand firm conduct. Unfortunately, we have little data to study conduct directly. Therefore, a basic exercise is to first estimate consumer behavior, then use the demand estimates to "reverse engineer" firm behavior and either test among competing theories of firm conduct or use a particular theory to simulate a counterfactual. For example, a researcher could estimate how consumers choose between different types of cars and use the estimates to compute the consumers' price sensitivity. Given this price sensitivity the researcher can compute the optimal markup implied by different theories of pricing and choose the theory that best fits observed data. In addition the researcher might also want to compute how the firms change their (pricing) behavior as a result of change in the environment, say due to a proposed merger or a change in regulation. See Bresnahan (1981) for an early example of this type of work, or Einav and Levin (2010) for a recent non-technical survey. Another reason IO economists are interested in consumer behavior is to measure consumer welfare. For example, we might want to evaluate the welfare effects of a proposed merger or the gain from the introduction of new goods.

Since consumer demand plays a key role in the above exercise, IO economists have spent significant time and effort in modeling and estimating demand, especially in industries with many differentiated products. In this paper, I will discuss some general lessons we learned from examining consumer behavior; and survey the main challenges and methods used to deal with these challenges. This paper is not a complete survey of demand modeling over the last couple of decades, and as such I leave out many developments and probably over emphasize IO related work. I try when possible to put the developments in IO within an historical context as well as relate to literature in related areas.

2 Some General Findings

Before surveying the methods it is useful to outline some general findings we have learned regarding consumer behavior. Offering these lessons up front helps explain some of the modeling choices emphasized in the literature. The two lessons are: (1) consumers view (even seemingly identical) products as differentiated and (2) consumers tastes are heterogenous.

2.1 Products are Differentiated

Economists tend to have strong priors regarding the relevance of differentiation, in many cases assuming that products are essentially identical. One of the key lessons learned from the data is that this is not true: almost all products are differentiated. It is easier to convince economists that some products are vertically differentiated. For example, at equal prices it is easy to claim that most (all?) consumers prefer a BMW to a Skoda. Differentiation arises, in equilibrium, because the price of the BMW will be higher and only some consumers are willing to pay the higher price.

Convincing economists that more narrowly defined products are horizontally differentiated is harder. For example, many will claim that Coke and Pepsi, or Post Raisin Bran and Kellogg Raisin Bran are essentially identical, that two supermarket chains are not differentiated in a meaningful way or that two American cars are not distinguishable. Consumers, however, tend to strongly disagree. When the price of one product declines we tend to see a decline in the sales of a competing product, but the decline is significantly less than what we would expect if the products were nearly homogenous. This finding is quite general and

is confirmed by many studies from numerous markets that vary by products, location and time, and use consumer level data or data aggregated at different levels.

There are many ways differentiation could arise. It could be due to inherent differences between products, information imperfections among consumers, marketing and advertising campaigns, or some sort of brand inertia. For some applications it is important to separate between these different explanations. Indeed, an interesting area of future research is to better understand the sources of this differentiation. However, from a more practical point of view, if one wants to explain consumer behavior this differentiation needs to be accounted for.

When working with data, one quickly learns that product attributes can explain some of the differentiation among products, but far from all of it. A store brand toasted oats cereal might have identical characteristics to General Mills Cheerios, yet even when Cheerios is priced much higher its sales are higher than the store brand. As we will see below, typically, this is accounted for by allowing for unobserved product level attributes, which will have important implications for how we estimate the model.

2.2   Consumers are Heterogenous

A second, somewhat related lesson, is the importance of consumer heterogeneity. Consumers are heterogenous in their tastes and in their income and as a result quite different in the choices they make. This is confirmed in market level data, but more importantly using consumer level choice data (for example, see Browning and Carro, 2007).

Interestingly, the heterogeneity in choice is only weakly correlated with standard consumer attributes. Income, education and family size obviously explain some dimensions of choice, but are far than enough to accurately predict consumer behavior. Unobserved heterogeneity is important to model in many cases.

3   Modeling Consumer Behavior

I now discuss how to model consumer demand in the presence of many differentiated products. I first outline the problem, then discuss some simple yet unsatisfactory, for IO purposes, solutions. The heart of this section is a discussion of the most commonly used models of demand.

3.1   The Problem

Suppose we are interested in estimating demand for J differentiated products. The most straight-forward approach to model consumer demand is to write down an aggregate demand system of the form

$$q = D(p, r, \varepsilon) \tag{1}$$

where q is a $J \times 1$ vector of quantities demanded, p is a $J \times 1$ vector of prices, r is a vector of exogenous variables, and ε is a $J \times 1$ vector of random shocks. Early work in demand estimation followed this approach, and the main modeling concern was to specify $D(\cdot)$ in a way that was both flexible and consistent with economic theory. Examples of resulting demand systems are the Linear Expenditure model (Stone, 1954), the Rotterdam model (Theil, 1965; and Barten 1966), the Translog model (Christensen, Jorgenson, and Lau, 1975), and the Almost Ideal Demand System (Deaton and Muellbauer, 1980a).

This approach, while intuitive, ends up being problematic in many cases considered in IO for several reasons.

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 linear demand system, $D(p, r, \varepsilon) = Ap + \varepsilon$, where A is $J \times J$ matrix of parameters. 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 options. Of course, with a more flexible functional form, the problem is even greater.

Second, in come cases the key interest is not aggregate demand, but a model of individual consumer behavior: 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. We should note that the mere presence of heterogeneity does not invalidate the approach of using an aggregate demand system. Under well specified conditions, that preferences are of the Gorman form (Gorman, 1959), we know that even with heterogeneity an aggregate demand system is well defined and can be treated as coming from a single representative consumer. The existence of heterogeneity does suggest that we should be careful in imposing

the restrictions of economic theory on the aggregate demand, since the conditions required for aggregation might not hold.

Third, and somewhat related, the aggregate representative consumer demand system does not easily allow for explicit parametrization of specific consumer behavior. For example, suppose we want to model demand for a storable good and account for the ability of consumers to store. A natural way to model this behavior is by an inventory model, where consumers make decisions based on their current inventory, storage costs, their expected future consumption needs and expected prices (see Hendel and Nevo, 2006b, for an example). The modeling exercise is much easier when we start with an explicit model of consumer behavior and aggregate to market level aggregate demand.

Fourth, this demand system does not easily allow us to predict the demand for new goods. As we will see below 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 product specific characteristics.

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 since we typically think that prices are correlated with the error terms and require an instrumental variable (IV) for each price. Finding a single IV is not easy, making it almost impossible to find enough IV that are both exogenous and will not generate moment conditions that are not nearly collinear.

3.2   Aggregation and Symmetry

Aggregation and symmetry are two, potentially easy, ways to solve some of the above issues, especially the dimensionality problem. Aggregation has a long history in demand analysis dating back at least to Gorman (1959). Symmetry assumptions were widely used in early theoretical models of product differentiation (Spence, 1976; Dixit and Stiglitiz, 1976). Both these approaches are very powerful but require strong assumptions that might be applicable in some cases but not in others.

One way to solve the dimensionality problem is to aggregate the individual products into aggregate commodities. In many cases aggregation might indeed make sense. In particular, if the researcher does not care about the substitution between the different products, only

the overall demand. For example, in some cases we might only want to know the demand for cars as a function of some average price. In this case we can estimate some version of equation (1) using only the total number of cars, but not the specific model.

Aggregation clearly has its advantages. The most important is that with more aggregation, possibly to a single aggregate, we can allow for flexible, even non-parametric, functional forms. But for many IO problems aggregating to the level of the industry misses the point: in many cases we care exactly about the substitution between the specific products. It is worth noting that almost all studies employ some level of aggregation. For example, most studies of the automobile market aggregate over various trims and option packages and define a product as a "model." So the real question is not whether to aggregate – we almost always do – but to what level and whether this aggregation solves the dimensionality problem.

The answer to the question of how much to aggregate depends on two things. First, what we are interested in. Obviously, if we care about substitution between products then aggregate demand cannot answer this question. However, even if we only care about total quantity, aggregation might be problematic. In order to aggregate we need to compute an average price, or price index. If prices of all the products we are aggregating over are highly correlated it is easy to compute this price (Hicks, 1936), but more generally computing the correct average price is more difficult. Without further assumptions one needs to know the substitution between the products – the exact thing we are trying to avoid having to estimate – in order to compute the correct price index. See Blundell and Stoker (2007) and references therein, for the various assumptions used in the literature in order to justify aggregation. So the second key to deciding on how much to aggregate has to do with the correlation of prices and the substitution between the products we are aggregating over: the more prices are correlated and the better substitutes the products the easier it is to compute the correct price to use.

An alternative way to solve the dimensionality problem is to impose symmetry across products. Theses type of models are used mostly in the trade and macro literature, as well as in the applied theory literature. The models tend to be easy to work with analytically, and can handle a large number of products. However, they cannot fit many patterns in micro data.

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, 1976),

presented here in its simplest form. Let the utility from consumption of the J products be given by

$$U(q_1, ..., q_J) = \left(\sum_{i=1}^{J} q_i^\rho \right)^{1/\rho}$$

where ρ is a constant parameter. This parametrization is quite popular because it combines a relatively simple functional form with a parameter that measures the taste for variety. For $\rho = 1$ we get linear preferences, or perfect substitution between products, while as $\rho \to \infty$ we get Leontief preferences, or perfect complements.

The demand of the representative consumer obtained from this utility function is

$$q_k = \frac{p_k^{-1/(1-\rho)}}{\sum_{i=1}^{J} p_i^{-\rho/(1-\rho)}} I \qquad i = 1, ..., J \tag{2}$$

where I is the income of the representative consumer.

Comparing equation (2) to equation (1) shows the power of the functional form assumption. Instead of having a number of parameters proportional to J^2, we have a single parameter to estimate, regardless of the number of products. We solved the dimensionality problem by imposing symmetry between the different products. To see this we note that the model implies

$$\frac{\partial q_i}{\partial p_j} \frac{p_j}{q_i} = \frac{\partial q_k}{\partial p_j} \frac{p_j}{q_k} \quad \text{for all } i, k, j$$

In words, the cross-price elasticities of i and k with respect to the price of j are restricted to be equal, regardless of how close a substitute the products really are. So while the functional form is convenient it imposes a very strong restriction on the demand system. The simplicity of the model and its analytic tractability make it a popular choice in theory and it is also heavily used in trade and in macro, but it is not appropriate to explain micro data and is essentially never used in empirical IO.

3.3   Most Commonly Used Demand Systems

The most commonly used demand systems in IO can be separated into two types: demand in product space and demand in characteristics space. The demand systems in product space continue to have a basic structure like that of equation (1), but solve the dimensionality

8

problem by assuming the utility is separable and therefore we can split the products into groups and estimate a flexible demand system within a group and between groups.

The demand systems in characteristics space solve the dimensionality problem by projecting the products onto a characteristics space. Within this class of models we will focus on discrete choice models. Recalling our first general lesson – the importance of product differentiation and the difficulty of capturing this differentiation with just product attributes – we will pay particular attention to the modeling of unobserved product attributes in the discrete choice model.

3.3.1 Separability

This class of models rely on an aggregate demand relation as in equation (1), but solve the dimensionality problem by dividing the products into smaller groups and allow for a flexible functional form within each group.

In order to formally motivate the split into groups, or segments, we would like to write the consumer's problem, of maximizing utility from consumption of the different products, as a sequence of separate but related decision problems. First, the consumer allocates expenditure to broad groups of products and then this expenditure is allocated to sub-groups of products, eventually allocated to a particular product. At each stage the allocation decision is a function of only that group total expenditure and prices of commodities in that group (or price indexes for the sub-groupings).

There are various conditions that will guarantee that the solution to this multi-stage process will equal the solution to the original consumer problem (see Deaton and Muellbauer, 1980a, chapter 5). One condition is to assume weak separability of preferences. Let $\widetilde{q}_1, \widetilde{q}_2, ..., \widetilde{q}_G$ be G subvectors of the vector $\widetilde{q} = (q_1, q_2, ..., q_J)$ such that each product is only in one group. Then the utility is weakly separable if

$$U(\widetilde{q}) = f\left(v_1(\widetilde{q}_1), v_2\left(\widetilde{q}_2\right), ... v_G\left(\widetilde{q}_G\right)\right)$$

where $f(\cdot)$ is some increasing function and $v_1, ..., v_G$ are the sub-utility functions associated with separate groups.

Weak separability is necessary and sufficient for the last stage of the multi-stage process; if a subset of products appears only in a separable sub-utility function, then the quantities

9

demanded of these products can always be written as only a function of group expenditures and prices of other products within the group.

In order to justify that the higher stages of the decision process – those that allocate expenditure between sub groups of products – further assumptions are needed. For example we can rely on indirect utility functions for each segment are of the Generalized Gorman Polar Form, and that the overall utility is separable additive in the sub-utilities (see Deaton and Muellbauer, 1980a, chapter 5, for a rigorous treatment).

The idea of multi-stage budgeting was originally developed for the estimation of broad categories of products such as food, clothing and shelter. Hausman, Leonard, and Zona (1994) and Hausman (1996) use the idea of multi-stage budgeting to construct a multi-level demand system for differentiated products. Their implementation is best illustrated by an example. Hausman, Leonard, and Zona (1994) estimate demand for beer and Hausman (1996) estimates demand for ready to eat cereal. Both papers have a similar structure with a category level demand as the highest level, a middle level that captures demand for specific segments (say family or kids cereal) and a lower level that represents demand for particular brands (Cheerios and Corn Flakes). Each level allows for a flexible functional form.

In particular, assume the data are for $j = 1, ..., J$ products in $t = 1, ..., T$ markets. In the lowest level they assume an Almost Ideal Demand System. The demand, or expenditure share, of product j in segment g in market t is given by

$$s_{jt} = \alpha_j + \beta_j \ln(y_{gt}/\pi_{gt}) + \sum_{k=1}^{J_g} \gamma_{jk} \ln(p_{kt}) + \varepsilon_{jt} \tag{3}$$

where, s_{jt} is the dollar sales share of product j out of total segment expenditure, y_{gt} is overall per capita segment expenditure, π_{gt} is the segment level price index, and p_{kt} is the price of product k in market t. This system defines a flexible functional form that can allow for a wide variety of substitution patterns within the segment. It has two additional advantages over other flexible demand systems (like the Rotterdam system or the Translog model): (1) it aggregates well over individuals; and (2) it is easy to impose (or test) theoretical restrictions, like adding-up, homogeneity of degree zero and symmetry (for details see Deaton and Muellbauer, 1980a).

The segment level price index, π_{gt}, is computed as either the Stone logarithmic price index

$$\pi_{gt} = \sum_{k=1}^{J_g} s_{kt} \ln(p_{kt}) \tag{4}$$

or the Deaton and Muellbauer exact price index

$$\pi_{gt} = \alpha_0 + \sum_{k=1}^{J_g} \alpha_k p_k + \frac{1}{2} \sum_{j=1}^{J_g} \sum_{k=1}^{J_g} \gamma_{kj} \ln(p_k) \ln(p_j). \tag{5}$$

The exact form of the price index does not seem to be important for the results (Deaton and Muellbauer, 1980a pg 316-317). If the latter is used the estimation is non-linear, while with the Stone index the estimation can be performed using linear methods.

The middle level of demand models the allocation between segments, and can be modeled using the Almost Ideal Demand System, in which case equation (3) is used with both expenditure shares and prices aggregated to a segment level (the prices are aggregated using either (4) or (5)). An alternative is the log-log equation used by Hausman, Leonard, and Zona (1994) and Hausman (1996):

$$\ln(q_{gt}) = \alpha_g + \beta_g \ln(Y_{Rt}) + \sum_{k=1}^{G} \delta_k \ln(\pi_{kt}) + \varepsilon_{gt}$$

where q_{gt} is the quantity sold of products in the segment g in market t, Y_{Rt} is total category (e.g., cereal) expenditure, and π_{kt} are the segment price indices (computed using either either (4) or (5)).

As we mentioned above, in order to be consistent with exact two-stage budgeting the segment level demand system needs to satisfy several conditions, which are not satisfied by commonly used demand systems. In some cases, the approach can still be justified as an approximate two stage budgeting approach. Although in practice these additional constraints are mostly ignored.

Finally, at the top level the demand for the category is specified as

$$\ln(Q_t) = \beta_0 + \beta_1 \ln(I_t) + \beta_2 \ln \pi_t + Z_t \delta + \varepsilon_t$$

where Q_t is the overall consumption of the category in market t, I_t is real income, π_t is the price index for the category and Z_t are variables that shift demand.

3.4   Models in Characteristics Space and Discrete Choice

Up to now we focused on demand in product space and looked for restrictions, through aggregation, symmetry or separability, to reduce the dimension of the problem. An alternative approach is to view a product as a collection of characteristics (Gorman, 1980, Lancester, 1966). The basic idea is somewhat similar to what we saw in the previous section: some products are better substitutes to each other than others. However, rather than separating the products into discrete segments we use the attributes of the products to derive their relative substitutability. The dimensionality problem is solved by making the relevant dimension the dimension of the characteristics, and not the number of products. A key issue to deal with is how to specify unobserved product attributes, which as we claimed in Section 2 are key to explaining the data. There are several ways to operationalize this approach, but the most popular, and the one I focus on here, is based on the discrete choice model.

A typical specification of the model starts with the indirect utility of consumer i from consuming product j in market t, $U(x_{jt}, \xi_{jt}, I_i - p_{jt}, \tau_i; \theta)$, which is a function of observed and unobserved (by the researcher) product characteristics, x_{jt} and ξ_{jt} respectively, income minus price, $I_i - p_{jt}$; individual characteristics, τ_i ; and unknown parameters, θ . Here I present a very simple, linear, utility model and discuss extensions later. Assume that the (conditional) indirect utility is

$$u_{ijt} = \alpha_i(I_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt} \tag{6}$$

where I_i is the income of consumer i, $x_{jt} = (x_{jt}^1, ..., x_{jt}^K)$ is a $1 \times K$ vector of observable characteristics of product j, and ε_{ijt} is a stochastic term. α_i is consumer's i marginal utility from income, β_i is $K \times 1$ vector of individual specific taste-coefficients.

An important part of this specification is the unobserved characteristic, ξ_{jt}. In many cases we might doubt the ability of observed characteristics to capture the essence of the product. For example, Hausman (1996, pg 229) comments that "it is difficult to conceive how I would describe Apple-Cinnamom Cheerios in terms of its attributes." The unobserved characteristic is meant to address these type of concerns. It captures unobserved attributes of the product, unquantifiable factors ("brand equity"), systematic shocks to demand, or unobserved promotional activity. An important lesson, which we stated in Section 2, is that this unobserved characteristic is essential to explain the data. For estimation, the existence

of ξ_{jt} implies that prices, as well as other choice variables, could be endogenous, if firms observe ξ_{jt} before making decisions.

The last part of the utility is the stochastic term, ε_{ijt}. This term is essential to explaining micro behavior: without it we cannot rationalize why consumers faced with the same choice set (and prices) make different choices. There are two types of interpretation of this shock, and the utility defined by equation (6). The first is that utility from a brand is deterministic, but the choice process itself is probabilistic (see for example, Tversky, 1972). The individual will not necessarily choose the alternative with the highest utility, rather has a positive probability of choosing each of the various options. Under this interpretation the ε's are not part of the utility, and only introduce randomness into the choice process, which is not taste-related. The second interpretation is that the true utility used by consumers to make choices is deterministic, but due to the researcher's inability to formulate individual behavior precisely an additional stochastic term is added. Thus, making utility stochastic from the researcher's point of view (see Manski 1977; and McFadden 1981, 1984). This is the interpretation followed in the economics literature and the one I follow here.

An interesting interplay is between ξ_{jt} and ε_{ijt}. At this point it might not be clear that we need both, in a way all the ξ_{jt} is doing is changing the mean of ε_{ijt}, by j and t. We will return to this point in Section 6.1, where we discuss consumer welfare and explore a model without ε_{ijt}.

While most studies take the indirect utility as the fundamental building block, we should note that we typically think of it as coming from a well specified utility maximization problem. Understanding the foundations underpinning the model is not just a formality, it imposes some restrictions on the functional form of the indirect utility. The formal derivation of the problem is beyond the scope of this paper, but let me demonstrate some of the issues. Suppose the consumer's preferences can be represented by a continuous utility function, $U(Q_0, Q_t)$, where Q_0 is the amount consumed of the numéraire and Q_t is the consumption of the "inside" good. This utility is maximized subject to a budget constraint. Conditional on choosing one of the J options for the inside good the conditional indirect utility can be written in some form like equation (6). For example, if utility is quasi linear, i.e., $U = f(Q_t) + Q_0$ with $f' > 0 > f''$, and $f(Q_t) = x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}$, then the conditional indirect utility will be given by equation (6). On the other hand, if the utility is Cobb-Douglas,

$U = Q_0^\alpha f(Q_t)^{1-\alpha}$, then the indirect utility will be given by

$$u_{ijt} = \alpha \ln(I_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}. \tag{7}$$

Note that while linearity of price might seem like a very special case, it is implied by quasi-linearity of preferences, which for many products seems like a reasonable assumption.

The derivation from an underlying utility function imposes at least two immediate restrictions. First, the term $I_i - p_{jt}$ should enter the utility and not p_{jt} alone. In the quasi-linear case, I_i, can be dropped without loss of generality since it will just shift all utility by a constant. More generally, however, this is not true. Second, if we believe that the underlying utility is given by $U(Q_0, Q_t)$ then indirect utility should be weakly separable in $I_i - p_{jt}$ and $f(Q_t)$. In other words, the interactions between the "price term" and product attributes are limited. For example, under this utility function it is not kosher to allow for a different price coefficient for each product. An alternative specification of utility defines it directly on the characteristics, $U(Q_0, x, \xi, \varepsilon)$, allows for interactions between the utility from the numéraire and the specific attributes. For some products this extra flexibility makes sense, but for others it does not. We return to this point when we discuss more flexible functional forms for the indirect utility.

The consumer-level taste parameters are modeled as

$$
\begin{aligned}
\alpha_i &= \alpha + \sum_{r=1}^{d} \pi_{1r} D_{ir} + \sigma_1 v_{i1}, \\
\beta_{ik} &= \beta_k + \sum_{r=1}^{d} \pi_{(k+1)r} D_{ir} + \sigma_{k+1} v_{i(k+1)} \quad \text{for } k = 1, ..., K
\end{aligned} \tag{8}
$$

where, $D_i = (D_{i1}, ..., D_{id})'$ is a $d \times 1$ vector of observed demographic variables, $v_i = (v_{i1}, ..., v_{i(K+1)})'$ is a vector of $K + 1$ unobserved consumer attributes. Π is a $(K + 1) \times d$ matrix of parameters and $\sigma = (\sigma_1, ..., \sigma_{K+1})$ is a vector of parameters. If we have individual level data then the demographics, D_i, are the individual attributes observed in this data. Sometimes we will not observe the demographics at the individual level but we will know their distribution, denoted by P_D. The joint distribution of $(v_{i1}, ..., v_{i(k+1)})$ is given by F_v, which is typically assumed to be standard normal.

The specification of the demand system is completed with the introduction of an "outside good": the consumers may decide not to purchase any of the brands. The indirect utility from this outside option is

$$u_{i0t} = \alpha_i I_i + \varepsilon_{i0t}$$

Let $\theta = (\alpha, \beta, \Pi, \sigma)$ denote the parameters of the model. Combining equations (6) and (8), and dropping the term I_i, which just shifts all utilities by a constant and therefore does not impact relative utilities or choice,

$$u_{ijt} = \delta_{jt}(x_t, p_t, \xi_t; \alpha, \beta) + \mu_{ijt}(x_t, p_t, D_i; \Pi, \sigma) + \varepsilon_{ijt}$$

where $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ is the mean utility across consumers. The variation around this mean is captured by two terms. The first,

$$\mu_{ijt} = -\left(\sum_{r=1}^{d}\pi_{1r}D_{ir} + \sigma_1 v_{i1}\right)p_{jt} + \sum_{k=1}^{K}\left(\sum_{r=1}^{d}\pi_{(k+1)r}D_{ir} + \sigma_{k+1}v_{i(k+1)}\right)x_{jt}^k$$

captures the interaction of consumer demographics and product attributes. The second, is the random term, ε_{ijt}. Before we have made any distributional assumptions the two terms are interchangeable.

For now we assume that consumers purchase one unit of the good, which gives the highest utility. We will later discuss how we can relax this assumption. Thus, the probability that a consumer of type (D_i, v_i) chooses option j is

$$s_{ijt} = s_{ijt}(x_t, \delta_t, p_t, D_i, v_i; \theta) = \int 1[u_{ijt} \geq u_{ikt}\ \forall l \mid x_t, \delta_t, p_t, D_i, v_i; \theta]dF_\varepsilon(\varepsilon). \qquad (9)$$

where $x_t = (x_{1t}, ..., x_{Jt})$, $\delta_t = (\delta_{1t}, ..., \delta_{Jt})$, $p_t = (p_{1t}, ..., p_{.Jt})$ and $1[A]$ is an indicator function that equals one if the event A is true. For estimation purposes we will integrate this probability over the unobserved consumer attributes, v_i, or even over all consumer attributes. (D_i, v_i), to get market shares

$$s_{jt} = s_{jt}(x_t, \delta_t, p_t; \theta) = \int s_{ijt}(x_t, \delta_t, p_t, D_i, v_i; \theta)dF_D(D)dF_v(v) \qquad (10)$$

In order to estimate the model, using either consumer or market level data, we make assumptions on the distribution of the (unobserved) individual attributes and compute this integral.

Different distributional assumptions will yield different models and have implications for the patterns of substitutions. Possibly the simplest assumptions we can make are that (1) $\Pi = 0$ and $\sigma = 0$, which implies $\beta_i = \beta$ and $\alpha_i = \alpha$ for all i; (2) ε_{ijt} are iid; and (3) ε_{ijt} are distributed according to a Type I extreme value distribution. These assumptions yield the (multinomial) Logit model and the market share of brand j in market t, is given by

$$s_{jt} = \frac{\exp\{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}\}}{1 + \sum_{k=1}^{J} \exp\{x_{kt}\beta - \alpha p_{kt} + \xi_{kt}\}} \tag{11}$$

This model is appealing due to its tractability, but it significantly restricts the substitution patterns. The price elasticities are

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

There are two problems with these elasticities. First, in most cases the market shares are small, so $\alpha(1 - s_{jt})$ is nearly constant and therefore the own-price elasticities are proportional to price. This implies that the lower the price, the lower the elasticity (in absolute value), and when plugged into a standard pricing model predicts a higher markup for the lower-priced brands. One question is whether this pattern is reasonable, but more importantly this pattern is a direct implication of the functional form. If, for example, indirect utility was a function of the logarithm of price, rather than price, then the implied elasticity would be roughly constant. In other words, the functional form directly determines the patterns of own price elasticity.

An additional problem, which has been stressed in the literature, is with the cross-price elasticities. We note that the cross price elasticity with respect to a change in the price of product k is that same for all products such that $j \neq k$. Essentially, what is happening is that when the price of k increases, some consumers will no longer view it as their top option and will substitute to their next option. Since the only heterogeneity across consumers is in the form of the iid ε_{ijt}, the consumers, who are no longer choosing k, value the other options like the average consumer and will choose that option at the same frequency as the market share. Hence, the percent change in the market share is constant.

In reality, we think that consumers who no longer choose option k are more likely, than the average consumer, to choose similar options. For example, consumers whose top option is a BMW are more likely to choose another luxury car as their second option. In order to capture this we need the variation around the mean utility to vary systematically across options. This can be done in one of two ways. First, we could generate the correlation by allowing ε_{ijt} to be correlated across j (i.e., relax assumption (2) above). Alternatively, we could generate the correlation by allowing for heterogeneity in the tastes (i.e., relax assumption (1) above). It is important to note, that assumption (3), of an extreme value distribution, allows us to obtain a closed form expression for the market shares, but otherwise it plays little role in driving the patterns of elasticities . The same issues are present if we assume other (iid) distributions. For example, if we assumed ε_{ijt} are distributed normal.

A simple model that attempts to deal with the problem of the cross price elasticities is the Nested Logit model. Continue to assume $\Pi = 0$ and $\sigma = 0$, and divide the products into mutually exclusive nests, $g = 1, ..., G$. Finally, let $\varepsilon_{ijt} = \lambda \varepsilon_{ig(j)t} + \varepsilon_{ijt}^1$, where ε_{ijt}^1 is an iid 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 $\varepsilon_{ig(j)t}$ (see Cardell, 1997) we get the Nested Logit model. Note, that if $\lambda = 0$ we are back to the Logit model. The Nested Logit model is a private case of the more general Generalized Extreme Value model (McFadden, 1978), which imposes correlation among the options through correlation in ε_{ijt}. In principle one could consider estimating an unrestricted variance-covariance 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. See Hausman and Wise (1978) for an application following this approach with a small number of options.

There have been several criticisms of these models. First, they do not deal with the problem with own price elasticities. Second, is the requirement for a-priori known segments. In principle the nesting structure can be tested. But in practice the tests are not very powerful.

A different solution to the problem with the elasticities is offered by the Mixed Logit or Random Coefficients Logit, as described by equations (6) and (8). An early version of this model was introduced by Boyd and Mellman (1980) and Cardell and Dunbar (1980). More recently, versions of the model were discussed in Berry, Levinsohn and Pakes (1995) and

McFadden and Train (2000). This model addresses both of the concerns with the elasticities by allowing for heterogeneity. We assume that ε_{ijt} are distributed iid according to a Type I extreme value distribution,[2] but generate correlation through μ_{ijt} by allowing heterogeneity in tastes for the product attributes to drive correlation.. So, for example, if "luxury" is an attribute of a car, then a consumer who likes one luxury car is more likely, then the average consumer to like another luxury car.

In this model the price elasticities are

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_i s_{ijt}(1 - s_{ijt}) dP_D(D) dP_v(v) & \text{if } j = k \\ \frac{p_{kt}}{s_{jt}} \int \alpha_i s_{ijt} s_{ikt} dP_D(D) dP_v(v) & \text{otherwise} \end{cases}$$

Now the own-price elasticity will not be driven solely by functional form. The partial derivative of the market shares will no longer be determined by a single parameter, α. Instead, each individual will have a different price sensitivity, which will be averaged to a product specific mean price sensitivity using the individual probabilities of purchase as weights. The price sensitivity will be different for different products. So if, for example, product j has lower prices and attracts more price sensitive consumers (i.e., they are more likely to purchase that product than the average consumer) its average price sensitivity will be higher, implying a lower equilibrium markup. Therefore, own price elasticities are not driven solely by functional form, but by the differences in the price sensitivity between consumers who purchase the various products.

The full model also 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). The correlation between μ_{ijt} and μ_{ikt} will induce correlation between s_{ijt} and s_{ikt}, and drive the substitution patterns. Indeed, McFadden and Train (2000) show that this model is general enough to approximate a wide class of choice problems.

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.

---

[2]In principle, we can also allow ε to be distributed according to a generalized exterme value distribution or other distributions, such as a normal distribution.

4 Econometrics

In this section I briefly discuss some of the main issues in estimating the demand models. The estimation of the Almost Ideal Demand System involves mostly standard linear and non-linear methods and therefore I will focus on the discrete choice model.

Data typically comes in one of two forms: consumer level and market level data. In both cases we see prices and observed attributes of all products. With market level data we see the total quantity sold of each product in a number of markets. We also observe the distribution of demographics, P_D, in each market. With consumer level data we see the match between consumers and their choices, as well as the demographics of consumers. In some cases we see multiple choices by the same consumer and in rare cases we have survey data of the second choice (i.e., what the consumers would have chosen if their top option was not available) Finally, in some cases we will not see consumer level data but might have some information by demographic group (e.g., the average age of consumers who purchased j).

Identification comes from seeing how choices change as the attributes (prices) change and as the available choices vary. For example, suppose initially we see a choice between three products, and then we see the choice when product 3 is no longer available. The change in the market share of products 1 and 2, tells us how close substitutes they are to product 3. The model then relates this to the relative importance of the various product characteristics. See Berry and Haile (2009a,b) for a formalization of this argument.

A key issue is the endogeneity of price (and other attributes). Endogeneity arises, just as it does in the text book examples of demand estimation, if there is correlation between price and the unobserved product characteristic, ξ_{jt}. This correlation can arise for different reasons, but the most natural is if the firms when setting prices know more about ξ than the econometrician (at the extreme case firms observe ξ when setting prices). Note, that this correlation can arise regardless of the level of aggregation. Therefore, a common claim that with consumer level data endogeneity is not a concern, is in general not correct.

To get an idea of how the model is estimated, and how we deal with endogeneity, suppose we have consumer level data on choices made by a sample of consumers $i = 1, ..., N$ in markets $t = 1, ..., T$, each with $j = 1, ..., J_t$ products. Let $y_{it} = j$ if the consumer choose product j in market t. Equation (9) gives us the probability of this choice. Suppose, for example, that ε_{ijt} are iid extreme value then

$$\Pr(y_{it} = j | x_t, \delta_t, p_t, D_i; \Pi, \sigma) = \int \frac{e^{\delta_{jt} + \mu_{ijt}}}{1 + \sum_{k=1}^{J_t} e^{\delta_{kt} + \mu_{ikt}}} dF_v(v).$$

The data give us this probability and allows us to estimate the parameters Π, and σ using simulated maximum likelihood, or simulated method of moments. The estimation will also recover a product-market specific constant

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}. \tag{12}$$

We can use the recovered product-market constants to estimate α and β, while dealing with the correlation of p_{jt} and ξ_{jt}, using standard methods, which we discuss below.

In some cases, however, consumer level data can help deal with the endogeneity problem. Suppose, for example, that prices vary by individual, yet ξ_{jt} does not. In such a case we could control for ξ_{jt} with a product-market level fixed effect. Similarly suppose that the indirect utility is given by equation (7), then the price coefficient, actually the coefficient on the $\log(I_i - p_{jt})$ term, can be identified from variation in income, I_i, across consumers, controlling for ξ_{jt} with a product-market level fixed effect. For any of these to work we need to observe multiple consumers in the same market purchasing the various products. However, with a large number of consumers we risk over fitting: as we average across many consumers we have no error left to explain why the model does not perfectly fit the data.

Suppose that instead of consumer level data we observe only market level shares of $j = 1, ..., J_t$ products in markets $t = 1, ..., T$. In order to use standard methods to deal with endogeneity we need to extract the error term, ξ_{jt}, from inside the non-linear share equation. The basic idea of the estimation is to invert the share equations, given by (10) in order to recover the mean utility given by (12). The inversion exists under general conditions as long as the products are substitutes (see Berry, Levinsohn and Pakes, 1995, or Berry and Haile, 2009b, for proof). Once we compute the mean utility we can write

$$\xi_{jt} = \delta_{jt}(s_t; \Pi, \sigma) - (x_{jt}\beta - \alpha p_{jt}).$$

Just as when we have consumer level data we can write the unobserved characteristics as a function of data and parameters.

Note that the market shares, observed in aggregate data, and the probability of purchase as a function demographics, observed in consumer level data, both play a similar role. The

key difference between the consumer level data and market level data is that with consumer level data we can see variation in the choice probability as a function of demographics holding the attributes, including the unobserved characteristics, fixed, while in the market level data we can see variation in the market shares as both (the distribution of) demographics and the unobserved attributes change. This difference allows for some additional flexibility in identification using consumer level data (Berry and Haile, 2009a), and is very helpful in estimation . Indeed, for estimation with aggregate data it is very useful to either have a very large number of markets, with varying demographics, or some other form of micro moments (i.e., purchase probabilities by demographic groups.)

Having derived an expression for the unobserved characteristic, ξ_{jt}, as a function of data and parameters, we can estimate the parameters of the model and deal with endogeneity. The basic idea is to find instrumental variables, z, such that

$$E(\xi_{jt}|z_{jt}) = 0. \tag{13}$$

The instrumental variables usually try to capture variation in cost across products and markets or variation in markups. Classical instruments for demand use variation in cost, such as input cost. Typically we have little cost information, especially by product, and therefore this approach is rarely used. Two exceptions are Nevo (2001), who uses measures of costs, and Villas-Boas (2007) who uses input costs interacted with product fixed effects.

Hausman (1996) and Nevo (2001) use an alternative approach that does not require direct measure of costs, instead relying on indirect measures. They use prices of the product in other markets. The assumption is that after controlling for common effects, the unobserved characteristics are independent across markets, while prices will be correlated across markets due to common marginal cost shocks. The assumption of independence across markets will be violated, for example, if unobserved promotional activities are correlated across markets.

An alternative approach is to generate instruments by relying on variation in markups. Berry Levinsohn and Pakes (1995), following on a similar idea in Bresnahan (1981, 1987), assume that $E(\xi_{jt}|x_{jt}) = 0$ and propose using functions of the characteristics of other products as instruments. The idea is that the markup varies with the degree of competition faced by the product, which is measured by the proximity in characteristics space to other products. The instruments are justified by assuming that x_{jt} are set without knowing ξ_{jt}. For instance, because they were set prior to the revelation of ξ_{jt}, and ξ_{jt} is not serially

correlated. Obviously, if ξ_{jt} is serially correlated the timing assumption is not sufficient to justify these instruments.

Another approach is to rely on panel data methods. In a simple form this just means assuming that ξ_{jt}, at least the part that is correlated with price, can be captured by a rich enough set of fixed effects. More recently, ideas from the dynamic panel data literature (Arellano and Bond, 1991, Blundell and Bond, 1998) have been used to motivate the use of characteristics as instruments. For example, we could assume that $\xi_{jt} = \rho \xi_{jt-1} + \eta_{jt}$, where $E(\eta_{jt}|x_{jt-1}) = 0.$Using this assumption $E(\xi_{jt} - \rho \xi_{jt-1}|x_{jt-1}) = 0$ is a valid moment condition

In many of these cases, the identifying assumptions required to justify the instruments have been questioned (for example, see the discussion by Bresnahan, 1996). For this reason, Nevo and Rosen (2009) build on the ideas of Manski and Pepper (2000), and explore using weaker identifying assumptions. Instead of relying on a moment equality as in (13) they build on a moment inequality and show that under certain conditions the parameters can be set identified. Applying this to the estimation of Logit demand they recover a reasonable, and potentially useful, set of parameters.

A separate issue is whether the instruments are "weak". This issue has rarely been explored in the IO literature, but could have important implications including problems with the standard errors and poor numerical performance.

The computation of the model typically follows the above steps of estimation (see Nevo, 2000b, for details and a computer code). Recently Dube et al (2009) offered an alternative computational method that bypasses the need for the inversion, instead solving a constrained optimization problem. Their method seems to work well, and speeds computation somewhat, especially if the number of market is not very large.

5   Comparing the Models

Having presented the most commonly used models, a natural question is how do they compare. Somewhat surprisingly there have been very few comparisons, either theoretical or empirical of the two main models. Judging by the academic literature, discrete choice models seem to be significantly more popular. In policy work, on the other hand, it seems like the preference has been for the simpler and maybe easier to estimate, multi-level demand system.

On a conceptual level the multi level demand system presented in Section 3.3.1 has some intuitive appeal: it is closer to classical demand models and seems to provide a flexible demand system within a segment. However, it has drawn some criticisms. First, the system requires classification of the products into segments. In many cases this segmentation is difficult to justify, but can be important for the bottom line. Supporters claim that different classifications can be tested against each other, but these tests are not very powerful and ultimately somewhat unconvincing. An approach that does not require weak separability, relying instead on latent separability that can be identified from the data, has been proposed by Blundell and Robin (2000).

Second, the derivation of the demand model, in principle, allows for aggregation of heterogenous preferences (assuming these preferences satisfy certain conditions). However, the derivation usually assumes that consumers consume positive amounts of all products. This is a reasonable assumption when the products are broad categories, but not with specific products. The typical consumer might consume more than a single brand, but rarely all brands. Little is known about the aggregation and approximation properties of the Almost Ideal Demand Model in this case. This is especially important since in many empirical applications the results are sensitive to whether or not we impose the restrictions of economic theory: adding up restrictions, symmetry and homogeneity. Whether or not we want to impose these conditions depends on whether we think the aggregate demand properly represents the demand of a representative consumer.

On the empirical side, the advantage of the multi level demand system is that it is simpler to estimate, requiring mostly linear estimation methods. Obviously, this saves on computational time, but maybe more importantly allows us to deal with measurement error in prices and shares. On the negative side, this system can typically be estimated only when there are a small, relatively constant across markets, number of products. And it requires a relatively large number of markets.

It also requires a large number of instrumental variables, which are hard to find in most applications. Indeed, the failure of the instruments is one of the explanations typically offered for a common pattern observed in empirical applications. Often products that we (strongly) believe are close substitutes end up being estimated as complements. For example, Hausman (1996) estimates that Kellogg Raisin Bran and Post Raisin Bran, have a negative (and statistically significant) cross price elasticity. This is not uncommon.

The discrete choice model we discussed in Section 3.4 is very popular in the academic IO literature but also draws a fair number of complaints. A common concern has to do with the assumption that consumers choose no more than one good. We know that many households own more than one car, that many of us buy more than one brand of cereal, and so forth. We note that even though consumers may buy more than one brand at a time, less actually consume more than one at a time. Therefore, the discreteness of choice can sometimes be defended by defining the choice period appropriately. In some cases this will still not be enough, in which case the researcher might view the model as an approximation, and then the question becomes if, and under what conditions, is it a reasonable approximation.

Empirically, the discrete choice model is often criticized when shares and prices are measured with error. Since it is a non-linear model the measurement error can cause significant biases. More importantly, in principle the model is flexible and can approximate many choice situations (McFadden and Train, 2000), but in reality the recovered distribution of heterogeneity might be quite restrictive and the model might be very close to the Logit model.

Huang, Rojas and Bass (2008) perform a Monte Carlo experiment comparing the performance of various models under different data generating processes. They generally find that a Logit model out performs the multi stage demand system. Their analysis is interesting but leaves many open questions, like a better understanding of the sources of bias and a study of the performance of additional demand structures.

6   Extensions of the Discrete Choice Model

In the academic IO literature the discrete choice model is by far the more popular choice for estimating demand. The basic model we presented has been extended in several ways. We briefly discuss some of these extensions here.

6.1   Consumer Welfare

One of the most common uses of demand models is to compute consumer welfare. This could either be the main motivation for the estimation (Trajtenberg, 1989, Nevo, 2003) or as a side to computing another counterfactual (for example, Nevo, 2000b, for mergers).

Computing consumer welfare using the discrete choice model is straightforward and relies on the inclusive value. McFadden (1978) defines the inclusive value (or social surplus) as the expected utility of a consumer, from several discrete options, prior to observing $(\varepsilon_{i0t}, ...\varepsilon_{iJt})$, knowing that the choice will be made to maximize utility after observing these shocks. When the idiosyncratic shocks ε_{ijt} are distributed i.i.d. extreme value, the inclusive value from a subset $A \subseteq \{1, 2, ..., J\}$ of the choice alternatives is defined as:

$$\omega_{iAt} = \ln \left(\sum_{j \in A} \exp \left\{ x_{jt} \ \beta_i - \alpha_i \ p_{jt} + \xi_{jt} \right\} \right) \tag{14}$$

When $\beta_i = \beta$ and $\alpha_i = \alpha$ the inclusive value captures the average utility in the population, averaging over the individual draws of ε, hence the term social surplus. When the utility is linear in price, as in equation (6), the inclusive value can be converted into a monetary equivalent by dividing by α_i. See McFadden (1981) and Small and Rosen (1981) for further details.

Petrin (2002) uses a discrete choice model to evaluate the welfare gains from the introduction of mini vans. He estimates a discrete choice model and uses it to compute a counterfactual of what the market would have looked like if the minivan were not introduced. He then uses the model again to compute the welfare in the two states of the world – the one observed and the counterfactual one – and attributes the difference to the introduction of the minivan. He claims that the Logit model, which does not allow for consumer heterogeneity, will overestimate the consumer gains. His 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. His solution, to try to reduce this effect, is to minimize the role of ε by relying more on random coefficients for heterogeneity. Berry and Pakes (2007) take this idea one step further and drop the epsilons all together in a model they call the pure characteristics demand model.[3]

As I argued above, allowing for heterogeneity in α_i and β_i is important, among other things, to generate reasonable elasticities. Indeed, allowing for heterogeneity can also have

---

[3] As I 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 characteristcs model and an alternative model.

an impact on the computation of welfare, but I think the source of the problem is slightly different then that identified by Petrin. The exercise Petrin performs has two steps: generating a counterfactual and then summarizing the counterfactual (and observed) prices and quantities into a welfare measure. Petrin identifies the second step as the source of the problem. I claim it is the first step that generates the problem, and I demonstrate this claim with the help of a classic example due to Debreu (1960) often called the "red-bus blue-bus example". Consider a market where consumers choose between driving their car to work or taking the red bus (for simplicity assume that working at home is not an option and that the decision of whether to work or not does not depend on the mode of transportation). Half the consumers choose a car and half choose the red bus. Now suppose we artificially introduce a new option: a blue bus. This option is artificial because consumers do not care about the color of the bus and in their eyes the red and blue buses are identical (suppose our consumers are color blind). Furthermore, suppose that prices are regulated, so they are not impacted by the introduction of the blue bus, and the frequency and quality of bus service is also not impacted. In reality, the introduction of this supposedly new option will result in an equilibrium where, as before, half the consumers choose a car, and the rest are split between the two color buses. Consumer welfare has not changed.

Now suppose we want to use the Logit model to analyze the consumer welfare generated by the introduction of the blue bus. Suppose we only observe data pre introduction of the blue bus and use it to estimate the model. Normalizing the mean utility from car, the outside good, to zero will yield $\delta_{car} = \delta_{(red)bus} = 0$, since $s_{car} = s_{(red)bus} = 0.5$, which implies an inclusive value of $\ln(e^0 + e^0) = \ln(2)$. Since the value of a blue bus is equal to the value of the red bus, i.e., $\delta_{red\_bus} = \delta_{blue\_bus} = 0$, if we use these estimates to simulate what the market would look like post introduction we will predict $s_{car} = s_{red\_bus} = s_{blue\_bus} = 1/3$, which implies an inclusive value of $\ln(3)$. In other words, we would predict a welfare gain when none was present.

Suppose we could eliminate the first step, of predicting the counterfactual market. This could be done if we observe the market post introduction. Given the above description the market share post-introduction will be $s_{car} = 0.5$ and $s_{red\_bus} = s_{blue\_bus} = 0.25$ implying $\delta_{car} = 0$ and $\delta_{red\_bus} = \delta_{blue\_bus} = \ln(0.5)$, and an inclusive value of $\ln(e^0 + 2 * e^{\ln(0.5)}) = \ln(2)$. So if we observed the correct market shares we would get the correct welfare estimate. Hence, in this example, and my claim is also more generally, the Logit model fails in the first step.

The reason this result holds more generally in the Logit model, and not just in this example, is that combining equation (11) with the inclusive value for all the options, given by equation (14) yields that the expected utility is $\ln(1/s_{0t})$. Since s_{0t} did not change in the observed data the Logit model predicted no welfare gain, but using the Logit model to generate the counterfactual market shares generated incorrect predictions. The Monte Carlo results in Berry and Pakes (2007) seem to provide a similar answer. They find that using the pure characteristics model matters for the estimated elasticities (and mean utilities) but not the welfare numbers. They conclude that "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."

Just to be clear, I am not claiming that the Logit would be a good model to use, just that we have to be clear what are its shortcomings. Furthermore, the difference between the Logit model and the Mixed Logit model in the change in welfare from period t to period $t-1$ is given by the difference between

$$\ln\left(\frac{1}{s_{0,t}}\right) - \ln\left(\frac{1}{s_{0,t-1}}\right) \quad \text{and} \quad \int\left[\ln\left(\frac{1}{s_{i,0,t}}\right) - \ln\left(\frac{1}{s_{i,0,t}}\right)\right]dP_D(D)dP_v(v).$$

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

6.2   Multiple choices

A common complaint about discrete choice models is that often they are applied to cases where choices are not discrete. For example, we might observe consumers buying several cans of soft drinks on a shopping trip, or households who own more than a single car. One way to rationalize the multiple choices is to assume that they are just aggregation over several choice instances. For example, a consumer shopping in a store is buying for a week. So assuming each day is a choice decision means that if the consumer bought 5 cans of soft drinks they decided to choose the outside option on two of the choice occasions. While providing a rationalization for the observed behavior this explanation is unappealing, in part because it assumes the choices across days are independent.

There are two potential issues to deal with when modeling multiple choices. First, the utility from product j might depend on whether product k is also chosen. Second, the choices could interact through a budget, or other, constraint.

Manski and Sherman (1980) study households choices of cars taking into account their current holdings. Their model accounts for the effect of past purchases on current decisions, but does not allow for simultaneous purchase of more than one option. Gentzkow (2007) looks at consumers choice between print and online newspapers, allowing for purchase of more than one option accounting for an interaction in the utility. In his model, consumers choose between the printed version of a newspaper, the online version, both, or neither. Thus, the choice is a discrete choice between bundles. Because the number of choices is small he is able to estimate the model using standard tools. However, for larger number of options a choice between bundles is not feasible to estimate this way. For example, for $J = 25$ there are $2^{25} = 33,554,432$ different bundles available.

Hendel (1999) studies a multi-discrete choice situation. In his case he observes firms simultaneously buying several brands of computers and several units of each brand, hence the nondiscreteness is in two dimensions. He models the choice of several brands as an aggregation over several tasks. The firm has several tasks to do. For each task there is an optimal brand, but the observed purchases are aggregation over several tasks. Note, that he does not allow for interaction in the utility from the different choices. The purchase of several units is explained by a decreasing marginal utility from quantity, hence there is interaction in this dimension.

Nevo, Rubinfeld and McCabe (2005) also study a multi choice problem. They examine the decision of libraries to subscribe to Economics and Business journals. There are over 150 possible journals a library can subscribe. All the libraries in their data subscribe to some subset of these journals, although the subsets are not nested (i.e., one could not model this is a choice of how many journals to purchase). They do not allow the utility from the journals to interact, but the interaction is through a budget constraint. Specifically, the journals are ranked by an index like that given in equation (6), and journals are purchased until a constraint is met.

6.3   Dynamics

The demand models discussed above are static. However, in many markets demand is dynamic in the sense that (a) consumers current decisions affect their future utility, or (b) consumers' current decisions depend on expectations about the evolution of future states. There is a long line of papers studying dynamic discrete choices. For example, Heckman (1981) and Flinn and Heckman (1982) study labor force dynamics where choices are dynamic in the sense that current decisions affect future states but consumers are not forward looking. Miller (1984), Wolpin (1984), Pakes (1986) and Rust (1987), study various decisions by economic agents using dynamic programming models of discrete choice. These methods have been applied widely.

In the context of demand for differentiated products, the exact effect of dynamics differs depending on the circumstances, and can be generated for different reasons. The literature has focused on several cases including storable products, durable products, habit formation, switching costs and learning. The key issue for this literature is how to write a model that accounts for all the products yet keeps the state space tractable. We summarize some of the key papers here. See Aguirregabiria and Nevo (2010) for a further review.

Consider storable products, if storage costs are not too large and current price is low relative to future prices (i.e., the product is on sale), there is an incentive for consumers to store the product and consume it in the future. Pesendorfer (2002) and Hendel and Nevo (2006a) present evidence that consumers indeed store when prices are low. Hendel and Nevo (2006b, and 2010) extend the above static models to allow for stoarability. They find that the static model overestimates the price elasticity and underestimates the cross price effects.

In the case of durable products, dynamics arise due to similar trade-offs. The existence of transaction costs in the resale market of durable goods (for example, because of adverse selection) implies that a consumer's decision today of whether or not to buy a durable good, and which product to buy, is costly to change in the future and, for that reason, it will impact her future utility. Therefore, when a consumer makes a purchase, she is influenced by her current holdings of the good and by her expectations about future prices and attributes of available products.

The impact of durable products on static estimation differ if we think there is repeat purchase or not. There are two problems with the standard static random coefficients discrete choice model if there are no repeat purchases (see Melnikov, 2000, and Conlon, 2010). First,

the distribution of the random coefficients is likely to change over time as some consumers purchase and exit the market. For example, if prices fall over time its likely that less price sensitive consumers purchase initially. Second, if consumers 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.

With repeat purchases the issues are a bit different (see Gowrisankaran and Rysman, 2009). First, the distribution of the consumers does not change, since consumers do not exit. However, consumers who previously purchased a product have a different value of no purchase since their alternative is to stay with their current product. Therefore, the problem with static estimation is that it does not account for the different value, across consumers and over time, of the outside option. Second, now when purchasing consumers realize that how long they hold onto the product is endogenous and therefore it changes their valuation of the options. For example, consumers might find it optimal to buy an inferior option – in the sense that it delivers lower flow utility – but replace quickly with a better/cheaper future option.

7   Concluding Comments

Demand estimation is at the heart of modern empirical IO. As a result IO economists have developed modeling and estimation methods, and certain norms about what is acceptable. As the IO community has grown some of these developments have been isolated from the rest of the profession. One interesting direction for future work is to explore more carefully connections with other areas of economics where models of consumer behavior have developed. These areas for the most part have developed separately from IO. See, for example, Blundell and Robin (2000), Lewbel (2001), Blundell, Browning and Crawford (2008), Blow, Browning and Crawford, (2008), and Lewbel, and Pendakur (2009).

Another direction for expansion and cross field fertilization is with other fields of applied micro. Recently there has been an increase in the use of various methods developed in IO. Hopefully, these methods will become common in other applied micro fields. The scope of applications of these methods is quite wide and as the set of applications increase interesting methodological issues are likely to arise. Furthermore, as IO economists work in areas

common to other applied micro fields some of the methods and concerns of these fields are likely to impact IO in general and studies of consumer behavior.

There is a long tradition in econometrics of using semi-parametric and non-parametric methods to estimate demand models as well as discrete choice models. The IO literature discussed has relied mostly on tightly specified parametric models focusing mainly on issues of endogeneity, consumer heterogeneity and product differentiation. Current non-parametric estimation can still not handle the dimensionality of the typical problem studied in IO. Future work, however, is likely to explore ways to relax some of the functional form assumptions currently made.

8 Literature Cited

Aguirregabiria, V and A Nevo (2010), "Recent Developments in Empirical Dynamic Models of Demand and Competition in Oligopoly Markets" mimeo.

Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economics Studies, 1991, 277-297.

Athey, Susan and Guido Imbens (2007), "Discrete Choice Models with Multiple Unobserved Choice Characteristics," International Economic Review, 48 (4), 1159-1192

Barten, A.P. (1966), Theorie en Empirie van een Volledig Stelsel van Vraagvergelijkingen, Doctoral dissertation, Rotterdam: University of Rotterdam.

Berry, Steven, and Phillip Haile. (2009a). "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers." Cowles Foundation Discussion Paper No. 1718.

Berry, Steven, and Phillip Haile. (2009b). "Identification in Differentiated Products Markets Using Market Level Data." Yale. Mimeo

Berry, Steven and Ariel Pakes (2007), "The Pure Characteristics Demand Model," International Economic Review, 48 (4).

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," Econometrica, 63, 841-890.

Blow, Laura, Martin Browning and Ian Crawford, (2008). "Revealed Preference Analysis of Characteristics Models," Review of Economic Studies, Blackwell Publishing, vol. 75(2), pages 371-389.

Blundell, R and S. Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," Journal of Econometrics

Blundell, R.,Martin Browning and Ian Crawford (2008) "Best nonparametric bounds on demand responses," Econometrica, 76(6), 1227-1262, November.

Blundell, R. and Jean-Marc Robin, (2000) "Latent Separability: Grouping Goods without Weak Separability," Econometrica, 68(1), pages 53-84, January.

Blundell, R and T. Stoker (2007),"Models of Aggregate Economic Relationships That Account for Heterogeneity," (with) in J. Heckman (ed.) Handbook of Econometrics, Chapter 68, pp 4609-4666.

Boyd, J. H., and Mellman, R. E., 1980, "The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis," Transportation Research, Part A, 14, pp. 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. (1987): "Competition and Collusion in the American Automobile Industry: The 1955 Price War," Journal of Industrial Economics, 35(4), 457-482.

Bresnahan, T. (1986), Comment on Hausman (1996) in T. Bresnahan and R. Gordon, eds., The Economics of New Goods, Studies in Income and Wealth Vol. 58, Chicago: National Bureau of Economic Research.

Browning, M. and J. Carro (2007), "Heterogeneity and microeconometric modeling", in Advances in Economics and Econometrics, volume 3, edited by Richard Blundell, Whitney Newey and Torsten Persson, Cambridge University Press, 2007.

Cardell, N.S. (1997), "Variance Components Structures for the Extreme Value and Logistic Distributions,"

Cardell, N.S. and F.C. Dunbar (1980), "Measuring the societal impacts of automobile downsizing", Transportation Research 14A, pp. 423–434.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau (1975), "Transcendental Logarithmic Utility Functions," American Economic Review, 65, 367-83.

Conlon, Chris, 2010, "A Dynamic Model of Costs and Margins in the LCD TV Industry", Yale mimeo.

Deaton, A., and J. Muellbauer (1980a), "An Almost Ideal Demand System," American Economic Review, 70, 312-326.

Deaton, A., and J. Muellbauer (1980b), Economics and Consumer Behavior, Cambridge University Press.

Deaton, A. (1986), "Demand Analysis," in Z. Griliches and M.D. Intriligator (eds.), Handbook of Econometrics, v. 3, pp. 1767-1839.

Debreu, G. (1960), "Review of R.D. Luce, Individual Choice Behavior: A Theoretical Analysis," American Economic Review, 50, 186-188.

Dixit, A., and J.E. Stiglitiz (1977), "Monopolistic Competition and Optimum Product Diversity," American Economic Review, 67, 297-308.

Dube, JP, Jeremy Fox, and C.-L. Su, (2009), "Improving the numerical performance of BLP static and dynamic demand estimation,"University of Chicago, mimeo.

Einav Liran and Jon Levin, (2010), "Empirical Industrial Organization: A Progress Report," Journal of Economics Perspectives, 24(2), Spring 2010, 145-162

Gentzkow, Matt, (2007) "Valuing New Goods in a Model with Complementarity: Online Newspapers," American Economic Review, June 713-44.

Gorman, W.M. (1959), "Separable Utility and Aggregation," Econometrica, 27, 469-81.

Gowrisankaran,G. and M. Rysman (2009): "Dynamics of Consumer Demand for New Durable Goods," manuscript. University of Arizona.

Hausman, J. (1996), "Valuation of New Goods Under Perfect and Imperfect Competition,"in T. Bresnahan and R. Gordon, eds., The Economics of New Goods, 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, 159-80.

Hausman, J., and D. Wise (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," Econometrica, 49, 403-26.

Heckman, J. and C. Flinn (1982), "New Methods for Analyzing Structural Models of Labor Force Dynamics," Journal of Econometrics, 18: 115-68

Hendel, Igal (1999), "Estimating Multiple Discrete Choice Models: An Application to Computerization Returns," Review of Economic Studies, April, 423-46.

Hendel, I., and A. Nevo (2006a), "Sales and Consumer Inventory," The RAND Journal of Economics, 37(3), 543-561.

Hendel, I., and A. Nevo (2006b), "Measuring the Implications of Sales and Consumer Inventory Behavior. Econometrica 74, 1637-1674.

Hendel, I. and A. Nevo (2010), "A Simple Model of Demand Anticipation," manuscript. Department of Economics. Northwestern University.

Hicks, J.R. (1936), Value and Capital, Oxford University Press.

Huang, D., C. Rojas and F. Bass (2008), "What Happens when Demand is Estimated with a Misspecified Model?" Journal of Industrial Economics, 56, 809-39.

Lewbel, A. (2001) "Demand Systems With and Without Errors," American Economic Review, 2001, 91, 611-618.

Lewbel, A. and Krishna Pendakur (2009), "Tricks With Hicks: The EASI Demand System," American Economic Review, June 2009, 99(3), 827-863.

Manski, C.F. (1977), "The Structure of Random Utility Models," Theory and Decision, 8, 229-254.

Manski, C. F., and J. V. Pepper (2000), "Monotone Instrumental Variables: With an Application to the Returns to Schooling,"Econometrica, 68(4), 997–1010.

Melnikov, O., 2000, Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market. Manuscript. Department of Economics, Yale University.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, eds., Frontiers of Econometrics, New York, Academic Press.

McFadden, D. (1978), "Modeling the Choice of Residential Location," in A. Karlgvist, et al., eds., Spatial Interaction Theory and Planning Models, Amsterdam: North-Holland.

McFadden, D. (1981), "Econometric Models of Probabilistic Choice," in C.F. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, Cambridge: MIT Press.

McFadden, D. (1984), "Econometric Analysis of Qualitative Response Models," in Z. Griliches and M. Intilligator, eds., Handbook of Econometrics, Volume III, Amsterdam: North-Holland.

McFadden, D. and Kenneth Train (2000) "Mixed MNL models for discrete response," Journal of Applied Econometrics, 15(5), pages 447-470.

Miller, R. (1984), "Job matching and occupational choice," Journal of Political Economy, Vol. 92, No. 6, pp. 1086-1120. 1984.

Moore, H.L. (1914), Economic Cycles: Their Law and Cause, New York: Macmillan.

Nevo, A (2000a), "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," Journal of Economics & Management Strategy, 9(4), 513-548, 2000.

Nevo, A (2000b), "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," The RAND Journal of Economics, 31(3), 395-421.

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, Daniel L. Rubinfeld and Mark McCabe (2005), "Academic Journal Pricing and the Demand of Libraries," American Economic Review, 447-452.

Nevo, Aviv, and Adam Rosen. 2009. "Identification with Imperfect Instruments." NBER Working Paper No. 14434.

Pakes, Ariel (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," Econometrica, 755-84.

Petrin, A. (2002) "Quantifying the Benefits of New Products: The Case of the Minivan," Journal of Political Economy, 705-29.

Pesendorfer, M. (2002): "Retail Sales: A Study of Pricing Behavior in Supermarkets," Journal of Business, 75(1), pages 33-66.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, 34-55.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," Econometrica, 55(5), 999-1033.

Schultz, H. (1938), The Theory and Measurement of Demand, Chicago: The University of Chicago Press.

Small, K. A., and H. S. Rosen, (1981), "Applied Welfare Analysis with Discrete Choice Models," Econometrica, 49, 105-30.

Spence, M. (1976), "Product Selection, Fixed Costs, and Monopolistic Competition," Review of Economic Studies, 43, 217-235.

Stigler, G.J. (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.

Theil, H. (1965), "The Information Approach to Demand Analysis," Econometrica, 6, 375-80.

Train, K. (2003), Discrete Choice Methods with Simulation. Cambridge, UK: Cambridge University Press.

Trajtenberg, M. (1989), "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," Journal of Political Economy, 97, 444-79.

Tversky, A. (1972), "Elimination by Aspects: A Theory of Choice," Psychological Review, 79, 281-299.

Villas-Boas, S. (2007), "Vertical Relationships Between Manufacturers and Retailers: Inference With Limited Data," The Review of Economic Studies, Vol. 74, 2, pp. 625-652

Wolpin, K. (1984), "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," Journal of Political Economy.