Lecture 1: NBERMetrics.

Ariel Pakes

July 17, 2012

We will be concerned with the analysis of demand systems for differentiated products and some of its implications (e.g.'s; for the analysis of markups and profits; for the construction of price indices, and for the interpretation of hedonic regression functions).

Why Are We Concerned With Demand Systems?

• The demand system provides much of the information required to analyze the incentive facing firms in a given market place. These incentives play a (if not the) major role in the determination of the profits to be earned from either

- different pricing decisions
- or alterantive investments (in the development of new products, or in advertising)

The response of prices and of product development to policy or environmental changes are typically the first two questions when analyzing either the historical or the likely future impacts of such changes. Advertising is a large part of firm investment and there is often a question of its social value and the extent to which we should regulate it. The answer to this question requires an answer to how advertising affects utility.

Note. There are other components determining the profitability of pricing and product development decisions, but typically none that we can get as good a handle on as the demand system. For example an analysis of the static response to a change in the environment would require also a cost function, and an equilibrium assumption. Cost data are typically proprietary, and we have made much less progress in analyzing the nature of competition underlying the equilibrium assumptions than we have in analyzing demand systems (indeed it would be easy to argue that we should be devoting our research efforts more towards understanding that aspect of the problem).

- Welfare analysis. As we shall see modern demand system analysis starts with individual specific utility functions and choices and then explicitly aggregates up to the market level when the aggregate demand system is needed for pricing or product placement decisions. It therefore provides an explicit framework for the analysis of the distribution of changes in utility that result from policy or environmental changes. These are used in several different ways.

  - Analyzing the consumer surplus gains to product introductions. These gains are a large part of the rationale for subsidizing private activity (e.g. R&D activity) and one might want to measure them (we come back to an example of this from Petrin's, *JPE* 2002, work, but it was preceeded by a simpler analysis of Trajtenburg, *JPE* 1989).

- Analyzing the impact of regulatory policies. These include pricing policies and the agencies decisions to either foster or limit the marketing of goods and services in regulated markets (regulatory delay, auctioning off the spectrum, ...). Many regulatory decisions are either motivated by non-market factors (e.g. we have decided to insure access to particular services to all members of the community), or are politically sensitive (partly because either the regulators or those who appointed them are elected). As a result to either evaluate them or consider their impacts we need estimates of the distributional implications.

- The construction of "Ideal price Indices". Should be the basis of CPI, PPI, ... Used for indexation of entitlement programs, tax brackets, government salaries, private contracts, measurement of aggregates.

Frameworks for the Analysis

There is really a two-way classification of differentiated product demand systems. There are

- representative agent (demand is derived from a single utility function), and
- heterogenous agents

models, and within each of these we could do the analysis in either

- product space, or
- characteristics space.

Multi-Product Systems: Product Space.

Demand systems in product space have a longer history than those in characteristic space, and partly because they were developed long ago, when they are applied to analyze market level data, they tend to use a representative agent framework. This is because the distributional assumptions needed to obtain a closed form expression for aggregate demand from a heterogenous agent demand system are extremely limiting, and computers were not powerful enough to sum up over a large number of individuals every time a new parameter vector had to be evaluated in the estimation algorithm. This situation has changed with the introduction of simulation estimators.

Digression; Aggregation and Simulation in Estimation. Used for prediction by McFadden, Reid, Travilte , and Johnson (1979, BART Project Report), and in estimation by Pakes, 1986.

Though we often do not have information on which consumer purchased what good, we do have the distribution of consumer characteristics in the market of interest (e.g. the CPS). Let z denote consumer characteristics, $F(z)$ be their distribution, p_j be the price of the analyzed, p_{-j} be the prices of competing goods.

We have a model for individual utility given (z, p_j, p_{-j}) which depends on a parameter vector to be estimated (θ) and generates that indiviudal's demand for good j,say $q(z, p_j, p_{-j}; \theta)$. Then aggregate demand is

$$D(p_j, p_{-j}; F, \theta) = \int_z q(z, p_j, p_{-j}; \theta) dF(z).$$

This may be hard to calculate exactly but it is easy enough to simulate an estimate of it. Take ns random draws of z from $F(\cdot)$, say $(z_1, \ldots, z_{ns})$, and define your estimate of $D(p_j, p_{-j}; F, \theta)$ to be

$$\hat{D}(p_j, p_{-j}; F, \theta) = \sum_{i=1}^{ns} q(z_i, p_j, p_{-j}; \theta).$$

If $E(\cdot)$ provides the expectation and $Var(\cdot)$ generates the variance from the simulation process

$$E[\hat{D}(p_j, p_{-j}; F, \theta)] = D(p_j, p_{-j}; F, \theta),$$

and

$$Var[\hat{D}(p_j, p_{-j}; F, \theta)] = ns^{-1} Var(q(z_i, p_j, p_{-j}; \theta)).$$

So our estimate is unbiased and can be made as precise as we like by increasing ns.

Product space and heterogenous agents. We will focus on heterogenous agent models in characteristic space for the reasons discussed below. However, at least in principle, demand systems in characteristic space are approximations to product space demand systems, and the approximations can be problematic[1]. On the other hand when it is feasible to use a product space model there are good reasons to use heterogeneous agent versions of it. This is because with heterogenous agents the researhcer

- can combine data from different markets and/or time periods in a sensible way, and in so doing add the information content of auxiliary data sets,
- and can get some idea of the entire distribution of responses.

The first point is important. Say we are interested in the impact of price on demand. Micro studies typically show that the price coefficient depends in an important way on income (or some more inclusive measure of wealth); lower income people care about price more. Consequently if the income distribution varies across the markets studied, we should expect the price coefficients to vary also, and if we did not allow for that it is not clear what our estimate applies to.

[1] I say in principle, because it may be impossible to add products every time a change in a product's characteristics occurs. In these cases researchers often aggregate products with slightly different characteristics into one product, and allow for new products when characteristics change a lot. So in effect they are using a "hybrid" model. It is also possible to explicitly introduces utility as a function both products and characteristics; for a recent example see Dubois,Griffith, and Nevo (2012, *Working Paper.*)

Further, even if we could chose markets with similar income distributions, we typically would not want to. Markets with similar characteristics should generate similar prices, so a data set with similar income distributions would have little price variance to use in estimation.

In fact as the table below indicates, there is quite a bit of variance in income distributions across local markets in the U.S. The table gives means and standard deviations of the fraction of the population in different income groups across U.S. counties. The standard deviations are quite large, especially in the higher income groups where most goods are directed.

Table I: Cross County Differences in Household Income*

Income Group (thousands)	Fraction of U.S. Population in Income Group	Distribution of Fraction Over Counties	
		Mean	Std. Dev.
0-20	0.226	0.289	0.104
20-35	0.194	0.225	0.035
35-50	0.164	0.174	0.028
50-75	0.193	0.175	0.045
75-100	0.101	0.072	0.033
100-125	0.052	0.030	0.020
125-150	0.025	0.013	0.011
150-200	0.022	0.010	0.010
200 +	0.024	0.012	0.010

* From Pakes (2004, *RIO*) "Common Sense and Simplicity in Empirical Industrial Organization".

Multi-product Systems in Product Space: Modelling Issues.

We need a model for the single agent and the remaining issues with proudct space are embodied in that. They are

- The "too many parameter problem". J goods implies on the order of J^2 parameters (even from linear or log-linear systems) and having on the order of 50 or more goods is not unusual
- Can not analyze demand for new goods prior to product introduction.

The Too Many Parameter Problem.

In order to make progress we have to aggregate over commodities. Two possibilities that do not help us are

- Hicks composite commodity theorem. $p_j^t = \theta_t p_j^0$. Then demand is only a function of θ and p_1, but we can not study substitution patterns, just expansion along a path.
- Dixit-Stiglitz (or CES) preferences. Usually something like $(\sum x_i^\rho)^{1/\rho}$. This rules out differential substitutions between goods, so it rules out any analysis of competition between goods. When we use it to analyze welfare, it is even more restrictive than using simple logits (since they have an idiosyncratic component), and you will see when we move to welfare that the simple logit has a lot of problems.

Gorman's Polar Forms.

Basic idea of multilevel budgeting is

- First allocate expenditures to groups
- Then allocate expenditures within the group.

If we use a (log) linear system and there are J goods K groups we get $J^2/K + K^2$ parameters vs $J^2 + J$ (modulus other restrictions on utility surface). Still large. Moreover multilevel budgeting can only be consistent if the utility function has two peculiar properties

- We need to be able to order the bundles of commodities within each group independent of the levels of purchases of the goods outside the group, or weak separability

  $$u(q_1, ..., q_J) = u[v_1(\vec{q}_1), ..., v_K(\vec{q}_K)].$$

  where $(q_1, ..., q_J) = (\vec{q}_1, ..., \vec{q}_K)$. which means that for i and j in different groups

  $$\frac{\partial q_i}{\partial p_j} \propto \frac{\partial q_i}{\partial x} \times \frac{\partial q_j}{\partial x}.$$

  (This implies that we determine all cross price elasticities between goods in different groups by a number of parameters equal to the number of groups).

- To be able to do the upper level allocation problem there must be an aggregate quantity and price index for each group which can be calculated without knowing the choices within the groups.

These two requirements lead to Gorman's (1968) polar forms (these are if and only if conditions);

- Weak separability and constant budget shares within a group (i.e. $\eta_{q,y} = 1$) (Weak separability is as above and it implies that we can do the lower level allocation, constant budget shares allows one to derive the price index for the group without knowing the "income" allocated to the group).
- Strong separability (additive), and budget shares need not go through the origin.

These two restrictions are (approximately) implemented in the Almost Ideal Demand Systems; see Deaton and Muelbauer (1980)

Grouping Procedures More Generally.

If you actually consider the intuitive basis for Gorman's conditions for a given grouping, they often seem a priori unreasonable. More generally, grouping is a movement to characteristics models where we group on characteristics. There are problems with it as we have to put products in mutually disjoint sets (which can mean lots of sets), and we have to be extremely careful when we "group" according to a continuous variable. For example if you group on price and there is a demand side residual correlated with price (a statement which we think is almost always correct, see the next lecture) you will induce a selection problem that is hard to handle. These problems disappear when you move to characteristic space (unless you group there also, like in Nested Logit, where they reappear). Empirically when grouping procedures are used, the groups chosen often depend on the topics of interest (which typically makes different results on the same industry non-comparable).

An Intro to Characteristic Space.

Basic Idea. Products are just bundles of characteristics and each individual's demand is just a function of the characteristics of the product (not all of which need be observed). This changes the "space" in which we analyze demand with the following consequences.

- The new space typically has a much smaller number of observable dimensions (the number of charactertistics) over which we have preferences. This circumvents the "too many parameter" problem. E.g. if we have k dimensions, and the distribution of preferences over those dimensions was say normal, then we would have k means and $k(k+1)/2$ covariances, to estimate. If there were no unobserved characteristics the $k(1+(k+1)/2)$ parameters would suffice to analyze own and cross price elasticities for all J goods, no matter how large J is.

- Now if we have the demand system estimated, and we know the proposed characteristics of a new good, we can, again at least in principle, analyze what the demand for the new good would be; at least conditional on the characteristics of competing goods when the new good is introduced.

Historical Note. This basic idea first appeared in the demand literature in Lancaster (1966, *Journal of Political Economy*), and in empirical work in McFadden (1973, "'Conditional logit analysis of qualitative choice behavior' in P. Zarembka, (ed.), *Frontiers in Econometrics*). There was also an extensive related literature on product placement in I.O. where some authors worked with vertical(e.g. Shaked and Sutton) and some with horizontal (e.g. Salop) characteristics. For more on this see Anderson, De Palma and Thisse (and the reading list has all these sites).

Modern. Transfer all this into an empirically useable theory of market demands and provide the necessary estimation strategies. Greatly facilitated by modern computers.

Product space problems in characteristic space analysis. The two problems that appear in product space re-appear in characteristic space, but in a modified form that, at least to some extent, can be dealt with.

- If there are lots of characteristics the too many parameter problem re-appears and now we need data on all these characteristics as well as price and quantity. Typically this is more of a problem for consumer goods than producer goods (which are often quite well defined by a small number of characteristics). The "solution" to this problem will be to introduce unobserved characteristics, and a way of dealing with them. This will require a solution to an "endogeneity" problem similar to the one you have seen in standard demand and suppy analysis, as the unobserved characteristic is expected to be correlated wth price.

- We might get a reasonable prediction for the demand for a new good whose characteristics are similar (or are in the span) of goods already marketed. However there is no information in the data on the demand for a characteristic bundle that is in a totally new part of the characteristic space. E.g. the closest thing to a laptop that had been available before the laptop was a desktop. The desktop had better screen resolution, more memory, and was faster. Moreover laptops' initial prices were higher than desktops; so if we would have relied on demand and prices for desktops to predict demand for laptops we would have predicted zero demand.

There are other problems in characteristic space of varying importance that we will get to, many of which are the subject of current research. Indeed just as in any other part of empirical economics these systems should be viewed as approximations to more complex processes. One chooses the best approximation to the problem at hand.

Utility in Characteristic Space.

We will start with the problem of a consumer choosing at most one of a finite set of goods. The model

- Defines products as bundles of characteristics.
- Assumes preferences are defined on these characteristics.
- Each consumer chooses the bundle that maximizes its utility. Consumers have different relative preferences (usually just marginal preferences) for different characteristics. $\Rightarrow$ different choices by different consumers.
- Aggregate demand. Sum over individual demands. So it depends on entire distribution of consumer preferences.

Formalities.

Utility of individual i for product j is given by

$$U_{i,j} = U(x_j, p_j, \nu_i; \theta), \quad \text{for} \quad j = 0, 1, 2, \ldots, J.$$

Note. Typically $j = 0$ is the "outside" good. This is a "fictional" good which accounts for people who do not buy any of the "inside" goods (or $j = 1, \ldots, J$). Sometimes it is modelled as the utility from income not spent on the inside goods. It is "outside" in the sense that the model assumes that its price does not respond to factors affecting the prices of the inside goods. Were we to consider a model without it

- we could not use the model to study aggregate demand,

- we would have to make a correction for studying a selected (rather than a random) sample of consumers (those who bought the good in question),
- it would be hard to analyze new goods, as presumably they would pull in some of the buyers who had purchased the outside good.

ν_i represents individual specific preferenc differences, x_j is a vector of product characteristics and p_j is the price (this is just another product characteristic, but one we will need to keep track of separately because it adjusts to equilibrium considerations). θ parameterizes the impact of preferences and product characteristics on utility. The product characteristics do not vary over consumers.

The subset of ν that lead to the choice of good j are

$$A_j(\theta) = \{\nu : U_{i,j} > U_{i,k}, \forall k\} \text{ (assumes no ties)},$$

and, if $f(\nu)$ is the distribution of preferences in the population of interest, the model's predictions for market shares are given by

$$s_j(x, p; \theta) = \int_{\nu \in A_j(\theta)} f(\nu) d(\nu),$$

where (x, p) without subscripts refer to the vector consisting of the characteristics of all J products (so the function must differ by j). Total demand is then $M s_j(x, p; \theta)$, where M is the number of households.

Details.

Our utility functions are "cardinal", i.e. the choices of *each* individual are invariant to affine transformations, i.e.

1. multiplication of utility by a person specific positive constant, and
2. addition to utility of any person specifc number.

This implies that we have two normalizations to make. In models where utility is (an individual specific) linear function of product characteristics, the normalizations we usually make are

- normalize the value of the outside good to zero. Equivalent to subtracting $U_{i,0}$ from all $U_{i,j}$.
- normalize the coefficient of one of the variables to *one* (in the logit case this is generally the i.i.d error term; see below).

The interpretation of the estimation results has to take these normalizations into account; e.g. when we estimate utility for choice j we are measuring the difference between its utility and that for choice zero, and so on.

Simple Examples.

- Pure Horizontal. In the Hotelling story the product characteristic is its location, person characteristic is its location. With quadratic transportation costs we have

$$U_{i,j} = \overline{u} + (y_i - p_j) + \theta(\delta_j - \nu_i)^2.$$

Versions of this model have been used extensively by theorists to gain intuition for product placement problems (see the Tirole readings). There is only two product characteristics in this model (price and product location), and the coefficient of one (here price) is normalized to one. Every consumer values a given location differently. When consumers disagree on the relative values of a characteristic we usually call the characteristic "horizontal". Here there is only one such characteristic (product location).

- Pure Vertical. Mussa-Rosen, Gabsewicz-Thisse, Shaked-Sutton, Bresnahan. Two characteristics: quality and price, and this time the normalization is on the quality coefficient (or, equivalently, the inverse of the price coefficient). The distinguishing feature here is that all consumers agree on the ranking of the characteristic (quality) of the different products. This defines a vertical characteristic.

  $$U_{i,j} = \overline{u} - \nu_i p_j + \delta_j, \quad \text{with} \quad \nu_i > 0.$$

  This is probably the most frequently used other model for analyzing product placement (see also Tirole and the references therein). When all consumers agree on the ranking of the characteristic across products we (like quality here) we call it a "vertical" characteristic. Price is usually considered a vertical characteristic.

- "Logit". Utiilty is mean quality plus idiosyncratic tastes for a product.

  $$U_{i,j} = \delta_j + \epsilon_{i,j}.$$

You have heard of the logit because it has tractible analytic properties. In praticular if ϵ distributes i.i.d. (over both products and individuals), and $F(\epsilon) = exp(-[exp(-\epsilon)])$, then there is a closed, analytic form for both

– the probability of the maximum choice conditioned on the vector of δ's and

– the expected utility conditional on δ.

The closed form for the probability of choice eases estimation, and the closed form for expected utility eases the subsequent analysis of welfare. Further these are the only known distributional assumption that has closed form expressions for these two magnitudes. For this reason it was used extensively in empirical work (especially before computers got good). Note that implicit in the distributional assumption is an assumption on the distribution of tastes, an assumption called the independence of irrelevant alternatives (or IIA) property. This is not very attractive for reasons we review below.

The Need to Go to the Generalizations.

It is useful to see what the restrictions the simple forms of these models imply, as they are the reasons we do not use them in empirical work.

The vertical model. $U_{i,j} = \delta_j - \nu_i p_j$, where δ_j is the quality of the good, and ν_i differs across individuals.

First, this specification implies that prices of goods must be ordered by their qualities; if one good has less quality than another but a higher price, no one would ever buy the lower quality good. So we order the products by their quality, with product J being the highest quality good. We can also order individuals by their ν values; individuals with lower ν values will buy higher priced, and hence higher quality, goods.

Now consider how demand responds to an increase in price of good j. There will be some ν values that induce consumers that would have been in good j to move to good $j+1$, the consumers who were almost indifferent between these two goods. Similarly some of the ν values would induce consumers who were at good j to move to $j-1$. However no consumer who choses any other good would change their choice; if the other good was the optimal choice before, it will certainly be optimal now.

So we have the implication that

- Cross price elasticities are only non-zero with neighboring goods; with neighbors defined by prices, and
- Own price elasticities only depend on the density of the distribution of consumers about the two indifference points.

Now consider a particular market, say the auto market. If we line cars up by price we might find that a Ford Explorer (which is an SUV) is the "neighbor" of a Mini Cooper S, with the Mini's neighbor being a GM SUV. The model would say that if the Explorer's price goes up none of the consumer's who substitute away from the Explorer would subsitute to the GM SUV, but a lot would substitute to the Cooper S. Similarly the density of consumer's often has little to do with what we think determines price elasticities (and hence in a Nash pricing model, markups). For example, we often think of symmetric or nearly symmetric densities in which case low and high priced goods might have the same densities about their indifference points, but we never expect them to have the same elasticities or markups.

These problems will occur whenever the vertical model is estimated; that is, if these properties are not generated by the estimates, you have a computer programming error.

The Pure Logit. I.e; $U_{i,j} = \delta_j + \epsilon_{i,j}$ with ϵ independent over products and agents. IIA problem. The distribution of a consumer's preferences over products other than the product it bought, does not depend on the product if bought. Intuitively this is problematic; we think that when a consumer switches out of a car, say because of a price rise, the consumer will switch to other cars with similar characteristics.

With the double exponential functional form for the distribution it gives us an analytic form for the probabilities

$$s_j(\theta) = \frac{exp[\delta_j - p_j]}{1 + \Sigma_q \, exp[\delta_q - p_q]}$$

$$s_0(\theta) = \frac{1}{1 + \Sigma_q \, exp[\delta_q - p_q]}$$

This implies the follwing.

- Cross price derivatives are $s_j s_k$. Two goods with the same shares have the same cross price elasticities with any other good regardless of that goods characteristics. So if both a

Yugo and a high end Mercedes increased their price by a thouand dollars, since they had about the same share[2] the model would predict that those who left the Yugo would be likely to switch to exactly the same cars that those who left the Mercedes did.

- Own price derivative $(\partial s/\partial p) = -s(1-s)$. Only depends on shares. Two goods with same share must have same markup in a single product firm "Nash in prices" equilibrium. So the model would also predict the same equilibrium markups for the Yugo and Mercedes.

Just as in the prior case, no data will ever change these implications of the two models. If your estimates do not satisfy them, there is a programing error.

Generalizations.

The generalizations below are designed to mitigate these problems, but when the generalizations are not used in a rich enough way in empirical work, one can see the implications of a lesser version of the problems in the empirical results.

- Generalization 1. ADT(1993) call this the "Ideal Type" model. Berry and Pakes (2007) call this the "Pure Characteristic" model. It combines and generalizes the vertical and the horizontal type models (consumers can agree on

[2]The Yugo was the lowest price car introduced in the states in the 1980-90's and it only lasted in the market for a few years; it had a low share despite a low price because it ws undersirable; the Mercedes had a low share because of a high price

the ordering of some characteristics and not others).

$$U_{i,j} = f(\nu_i, p_j) + \sum_k \sum_r x_{j,k} \nu_{i,r} \theta_{k,r},$$

where again we have distinguished price from the other product characteristics as we often want to treat price somewhat differently. A special case of this (after appropriate normalizations) is the "ideal type" model

$$U_{i,j} = f(\nu_i, p_j) + \delta_j + \sum_k \alpha_k (x_{j,k} - \nu_{i,k})^2.$$

- Generalization 2. BLP(1995)

$$U_{i,j} = f(\nu_i, p_j) + \sum_k x_{j,k} \nu_{r,k} \theta_{k,r} + \epsilon_{i,j}.$$

BLP vs The Pure Characteristic Model. The only difference between BLP and the pure characteristic model is the addition of the logit errors (the $\epsilon_{i,j}$) to the utility of each choice. As a result, from the point of view of modelling shares, it can be shown that the pure characteristics is a a limiting case of BLP; but it is not a special case and the difference matters[3]. In particular it can only be approached if parameters grow without bound. This is a case computers have difficulty handling, and Monte Carlo evidence indicates that when we use the BLP specification to approximate a data set which abides by the pure characteristic model, it can do poorly.

This could be problematic, because the logit errors are there for computatinal convenience rather than for their likely ability

[3]It is easiest to see this if you rewrite the BLP utility equation with a different normalization. Instead of normalizing the coefficient of ϵ to one, let it be $\sigma_\epsilon > 0$. Now divide by σ_ϵ. As $\sigma_\epsilon \to 0$, the shares from the pure characteristics model converge to those from BLP.

to mimic real-world phenomena. Disturbances in the utility are usually thought of as emanating from either missing product or missing consumer characteristics, and either of these would generate correlation over products for a given individual. As a result, we might like a BLP-type specification that nests the pure-characteristic model.

When might the logit errors cause problems? In markets with a lot of competing goods, the BLP specification does not allow for cross-price elasticities which are too large. This is because the independence of the logit errors implies that there are always going to be some consumers who like each good a lot more than other goods, and this bounds markups and cross price-elasticities. There is also the question of the influence of these errors on welfare estimates, which we will come back to in another lecture.

These are reasons to prefer the pure characteristics model. However, for reasons to be described below, it is harder (though not impossible) to compute, and so far the recently introduced computational techniques which you will be introduced to in the subsequent lecture, have not helped much on this problem. This is the reason that you have seen more of BLP than the pure characteristic model. Despite this its intuitive appeal and increased computer power have induced a number of papers using the pure characteristic model; often in markets where there are one or two dominant vertical characteristics (like computer chips, see Nosko, 2011). We come back to this in our discussion of computation.