Demand Systems in Industrial Organization: *

Allan Collard-Wexler

[Econ 890-02: Fall 2019]

Contents

*These notes draw from a variety of sources: in particular Ariel Pakes' lecture notes, and from (co-teaching with) John Asker at NYU Stern, and Robin Lee's notes at Harvard.

1 Overview

Demand systems often form the bedrock upon which empirical work in industrial organization rests. The next few lectures aim to introduce you to the different ways empirical researchers have approached the issue of demand estimation in the applied contexts that we typical confront as IO economists. I will start by briefly overviewing the types of research questions and various instances in which demand estimation is useful, and the core problems we face when estimating demand.

We will begin with a basic overview of homogeneous product market competition (with which you should be familiar), and an overview of estimation in these markets. We will then move to models of differentiated product demand systems. I will review basic theory and standard data forms, after which I will go on to talk about the standard approaches to demand estimation and their advantages and disadvantages. All these approaches try to deal with the problem of estimating demand when we are in a market with many, differentiated goods. Specific papers will be used to illustrate the techniques once they have been discussed.

I will expect you to remember your basic econometrics, particularly the standard endogeneity problem of estimating demand (see Working 1927 or the treatment in standard econometrics texts, e.g. Hayashi 2000 in Ch 3).

There has been an explosion in the sophistication of technique used in demand estimation the last decade, due to a combination of advances in econometric technique, computation and data availability.

1.1 Why spend time on Demand Systems?

Many questions in IO require understanding how consumers choose among various goods and services as a function of market and individual characteristics. Though properly estimating a demand

TABLE I

Year	(1) Auto Production[a]	(2) Real Auto Price-CPI[b]	(3) % Change Auto Price-Cagan[c]	(4) Auto Sales[d]	(5) Auto Quantity Index[e]
1953	6.13	1.01	NA	14.5	86.8
1954	5.51	0.99	NA	13.9	84.9
1955	7.94	0.95	−2.5	18.4	117.2
1956	5.80	0.97	6.3	15.7	97.9
1957	6.12	0.98	6.1	16.2	100.0

Notes: [a] Millions of units over the model year. [Source: *Automotive News.*]
[b] (CPI New automobile component)/CPI. [Source: *Handbook of Labor Statistics.*]
[c] Adjusted for quality change. [See Cagan (1971), especially pp. 232–3.]
[d] Auto output in constant dollars, *QIV* of previous year through *QIII* of named year, in billions of 1957 dollars. [Source: *National Income and Product Accounts.*]
[e] (4)/(2), normalized so 1957 = 100.

TABLE II

Year	(6) Per Capita Disposable Personal Income[f]	(7) Interest Rate[g]	(8) Durables Expenditures (Non-Auto)[h]	(9) Automakers Profits[i]
1953	1623	1.9	14.5	2.58
1954	1609	0.9	14.5	2.25
1955	1659	1.7	16.1	3.91
1956	1717	2.6	17.1	2.21
1957	1732	3.2	17.0	2.38

Notes: [f] Billions of 1957 dollars, *QIV* of previous year through *QIII* of named year. [Source: *National Income and Product Accounts.*]
[g] Three-month *T*-bill rate. [Source: *Statistical Abstract.*]
[h] Durables component of consumer expenditures minus component for automobiles and parts,

system in its own right may be an objective of interest, demand systems (and their underlying parameters) are more often than not used as an input into answering other, perhaps larger, questions. E.g., they are often used as providing the incentives for examining firm behavior (pricing, investment, product introduction, entry/exit, etc...), or computing consumer welfare from a policy change. For example...

- Infer firm conduct: sometimes it is difficult to observe/measure firm conduct directly, but we might be able to test certain theories by using consumer demand estimates to infer firm behavior.

  - **Example: Bresnahan 1987 Competition and Collusion in 1950s Auto Market**

    Bresnahan wanted to examine the hypothesis that the dramatic increase in quantity (45% greater than in two surrounding years) and decrease in the price of Autos in 1955 was due to the temporary breakdown of a collusive agreement. Unlikely to be demand shock: "any explanation of all of the 1955 events from the demand side will need to be fairly fancy."

    His idea was to assume that marginal costs were not varying and then ask whether the relationship between pricing and demand elasticities changed in a manner consistent with a shift from collusion to oligopolistic pricing.

He exploits data on P and Q for different makes of automobiles. He has about 85 models over 3 years. The "magic" in these approaches is using demand data *combined with* an equilibrium assumption on firm conduct to back out marginal costs, *without using any cost data*. We'll come back to this later.

- Welfare impacts: to conduct welfare calculations subsequent to some market change brought about by, say, policy intervention, product introduction, or etc., one needs a well specified demand system. It allows us to quantify the "Value of Innovation": e.g., compute consumer surplus from the introduction of a new good (e.g., minivans, CAT scans) with similar "characteristics" of existing ones.

- Determinants of Innovation: with a demand system, a researcher can compute predicted markups for a given good; consequently, one will understand the types of products a firm will want to produce (e.g., minivans or SUV's, cancer drugs instead of malaria treatments). Demand systems, in other words, help us measure the incentives for investing in new goods.

- Usually demand is important to think about various forms of comparative statics: common ones for IO researchers include pre and post merger pricing, tax incidence, monopoly vs duopoly pricing, effect of predatory pricing policies, impact of new product introductions, etc.

- In IO and Marketing, there is considerable work on advertising which usually involves some demand estimation. This about policy questions of direct-to-consumer drug adverting, or advertising as a barrier to entry. Furthermore, carefully specified demand systems can assist with decomposing the mechanisms or channels through which various advertising (and other) effects work. E.g., persuasive vs. informative advertising.

- Understanding the cross-price elasticities of good is often crucial to "preliminary" issues in policy work, such as market definition in antitrust cases. Also, they inform determinants of market power: should we allow two firms to merge? Is there collusion going on in this industry (unusually large markups)? Cross-price elasticities are one input into this equation. (We will talk a bit (later) about the myriad antitrust applications of demand models. Note that this is the largest consumer of Ph.D's in Empirical I.O. by a long shot!)

- The tools used in demand estimation are starting to be applied in a variety of other contexts (e.g., political economy, development, education, health...) to confront empirical issues, of there is likely to be some intellectual arbitrage for your future research.

2 Brief Theory Review

Before diving into estimation, it is useful to begin with a few basic theoretical models of quantity and price determination in markets. Predictions in these models will depend on characteristics of demand, which we will in turn discuss ways of estimating.

2.1 Homogenous Goods.

We begin with simple homogenous good markets provided by an oligopolistic set of firms competing in quantities (Cournot) or price setting (Bertrand).

2.1.1 The Cournot Model.

Assume that:

- firms choose quantities and
- a Nash equilibrium in quantities results.

where:

- $p(Q)$ is the inverse demand curve
- $Q = \sum_j q_j$, where q_j is the output of firm j,
- $C_j(\cdot)$ and $mc_j(\cdot)$ are the total and marginal cost functions.

Note. We now work explicitly with different mc curves for each firm, but sufficing with an "aggregate" approximation to the demand curve. Provided that is a good enough approximation, it will not hurt the implications of the model we do investigate (like price, efficiency in production)... However it will not allow us to get at the implications of the equilibrium on the distribution of consumer utilities and welfare.

Admitting heterogeneous marginal costs, however, will enable us to study "distributional" implications of the equilibrium on the supply side; for e.g., how efficient is the allocation compared to the least cost allocation of production for a given total quantity. This is a question which underlies many regulatory issues (e.g. increasing returns or large fixed cost justifications for monopoly and the output allocations that are generated by the ensuing regulations...)
The profit function for firm j is

$$\pi_j(q_j) = p(Q)q_j - C_j(q_j).$$

Assuming differentiability, and that all firms produce *positive* quantities, the f.o.c. for Cournot-Nash equilibrium provide a system of J equations in J unknowns

$$p(Q) + q_j \frac{\partial p}{\partial Q} - mc_j(q_j) = 0 \; \forall j$$

Of course even if $q > 0$, this is only a necessary condition for an equilibrium. For the solution to this system to truly be an equilibrium the choice of q_j must also satisfy the second order condition for each agent

$$2\frac{\partial p}{\partial Q} + q_j \frac{\partial^2 p}{\partial^2 Q} - \frac{\partial mc_j}{\partial q_j} < 0.$$

Jointly sufficient conditions for this are that:

- marginal revenue slopes down, and
- marginal cost slopes up.

Of course one can get by with weaker conditions.

Question. Assume m.c. is constant and that the demand curve has a constant elasticity. What then is a sufficient condition for the s.o.c. to be satisfied?

Note that both the necessary and sufficient conditions change when either;

- there is a *fixed* or *sunk* cost of production for either of these may provide a reason for a plant to not close down but still not produce (i.e. to "mothball" the plant) even if there is a cost to mothballing (a cost of maintaining the plant and the cost of re-entering in a future year). I.e in these cases we cannot assume *a priori*, that $q > 0$, for all active plants and we have to take account of corner solutions.

- there are capacity constraints, in which case the derivative of the cost function doesn't exist at the capacity constraint. We cannot perform the experiment of increasing q above capacity, so we cannot get the derivative from the right. We can get the derivative from the left and what we know is that for us to produce at the capacity constraint it must be positive.

These two cases are cases when the cost function is non-differentiable at a point of interest (either at zero, or at the capacity constraint), and any other nondifferentiability will cause related problems, it is just that these two are often relevant in applied work. We come back to them presently, but first we consider the implications of this model when the cost functions are "sufficiently smooth".

What are the *efficiency* implications of the allocation?

- Among the set of interior firms, larger firms have lower marginal cost at the produced quantities. So if you believe your data is being generated by a homogeneous product quantity setting model, and you believe that all firms are at an interior equilibrium point, then provided your data show a large variance in output across firms (and recall that almost all data sets do), then you must believe that some firms have much lower marginal costs and much higher markups at the outputs they produce than do other firms. This in turn implies that from the point of view of allocating a fixed amount of production we can do *better* than the market by allocating *more output* to larger firms (a "social planner" would equate marginal costs).

- Compare the "smooth" allocation to the situation in either competition or monopoly. We think that the allocation of *fixed* amount of output would be worse if that output were allocated according to a Nash equilibria than it would be if say, a multiplant monopolist were allocating the output. Where the gains from competition come in is not from the efficiency of the allocation of a given amount of output, but rather from the quantity of output produced (which we expect to be larger in competition in a Nash equilibria).

- Thus consolidation of the industry (through mergers or buyouts) is likely to have two effects;
  - there is an additional incentive to increase price, thereby decreasing consumer surplus
  - it is likely to improve the productivity or efficiency of the productions allocation, which, all else equal, increases producer surplus.

  You can find functional forms where either the first or second dominates in calculating total surplus. Indeed you can find functional forms where the gains in productivity actually are enough to overcome the incentive to increase price, and price will fall.

- The price term is often exhibited as a percentage markup, which in turn equals our Lerner index; i.e.

$$p(Q) = mc_j - q_j \frac{\partial p}{\partial Q} \Rightarrow \; L_j = (p - mc_j)/p = s_j/\eta$$

where s_j is the market share, and η is the absolute value of the demand elasticity. Consequently the average price cost-margin is just

$$\sum_j s_j (p - mc_j)/p = \sum_j s_j^2/\eta = H/\eta,$$

where H is the Herfindahl index of concentration.

- If there is a unique equilibrium, they define a best reply function for each firm to each vector of rivals outputs; these are generally written as

$$q_j = r_j(q_{-j}).$$

However even if there is not a unique equilibrium, once the competitors play is fixed the response of the firm in question is "generically" unique. This fact will be quite useful when we move to empirical work. We come back to other properties of the reaction function below.

2.1.2 Homogeneous Products Bertrand Competition

The next analysis of the homogeneous product model was by Bertrand. Noting that firm's often seem to set prices, he assumed price competition rather than quantity competition. The model then has properties which do not seem to square with reality. As a result it is a model which is not really used in empirical work, except in modified form. Still it was a useful theoretical device as it generated both questions and insights that ended up pointing the way towards a series of developments. Moreover it is used repeatedly in applied theory papers because of the simplicity of the equilibria it generates.

The standard Bertrand homogeneous product model with two agents (and it will become obvious how to generalize to a larger number of agents) has $D(p)$ being total demand, and if $D_1(p_1, p_2)$ is the demand for the first firm's product given both prices we have it equal to

$$= D(p_1) \text{ if } p_1 < p_2,$$

$$= (1/2)D(p_1) \text{ if } p_1 = p_2, \text{ and}$$

$$= 0 \text{ if } p_1 > p_2.$$

If there are constant costs,

$$\pi_1(p_1, p_2) = D_1(p_1, p_2)(p_1 - c).$$

Note that viewed as a function of p_1, $D_1(\cdot)$ and hence $\pi_1(\cdot)$ has a jump in it at $p_1 = p_2$. Thus one can not analyze necessary conditions as "zero derivative conditions", and to analyze Nash equilibria behavior we have to just compute profits at each p_1 given p_2 and figure out which price maximizes profits. If you run through the reasoning here you will see that the only Nash equilibrium is

$$p_1 = p_2 = c.$$

(It is easy to show it is an equilibrium; to show that nothing else can be let one firm have a price higher than c and show that the second firm's optimal response is to cut price....).

There are strong implications of this model.

- First it says that provided there are two firms in the industry with the lowest cost, the industry will act as if it is a "price taker" in the sense that the equilibrium is $p = mc$ (at that cost).
- With different, but constant costs, there is a single producer but it produces at a price just under the cost of *the second lowest cost producer*. So now we have production efficiency, but we are producing under the total surplus maximizing quantity. The extent of the deviation between price and minimal marginal cost depends on how close the most efficient rival's cost is to the efficient cost.

The results are *suspect* for many reasons.

- First and foremost, if there are any sunk or any fixed costs, this industry will never see entry (or production), by more than one plant. Thus the industry should be a monopoly, and the monopolist *should not* charge $p = mc$ for it maximizes current profit by setting the monopoly price, and its ability to deter entry depends only on its costs; i.e. provided the firm's costs can be revealed in a verifiable way and there are no collusive possibilities, there will be no entry no matter what price it sets. *So we ought not ever see this equilibrium.*
- Similarly we often think we observe small differences in prices existing without one firm dropping to zero demand. That is the discontinuity in the demand curve seems not to be true empirically.

There are a number of ways of getting around these unrealistic predictions, and we list some of them here. Note however that they do not necessarily get around all the problems, at least not without further assumptions[1].

- Differentiated Product Models. Then goods marketed are not perfect substitutes for one another (at least not to all consumers), so when one decreases its price below its competitor's price not all consumers jump to it. We will discuss this in quite a lot of detail, as it has become the dominant form of analysis in empirical work on consumer goods markets, and to a lesser extent on producer goods markets also (many of these have relatively homogenous goods, think industrial chemicals, but location and the costs of transportation are differentiating factors).
- Prices have rigidities (i.e. they cannot be adjusted too quickly). Simplest case; two firms, must hold price fixed for two periods, firms move in alternating periods. We are then in a *dynamic game*, where we make a price choice for this period and it determines both my price, and indirectly, the price of my competitor in the next period. We come back to this when we discuss dynamic games where we show that there can be more than one equilibria; one we will develop is an Edgeworth cycle - "wars of attrition", another is kinked demand curves; these solutions are discussed in the Maskin Tirole articles.

[1] E.g. it is easy enough to write down differentiated product models wherein a Bertrand equilibrium does not exist.

- Collusion (richer strategy spaces). Begins in the framework of "repeated" games; Green and Porter, Abreu Pearce and Stachetti, ... We set a price higher than mc and we enforce this price by punishing a firm from deviating from this price in future periods. See our analysis of collusion below.

- Capacity constraints, or more generally, increasing marginal cost. We discuss this briefly here.

Two Firms w/ Capacity Constraints, Competiting a la Bertrand. Consider the following solution to a two-firm model with capacity constraints. Of course if the capacities are never binding (capacity greater than $D(c)$ for both firms), then the capacity constraints will not change the nature of the equilibrium. So assume that the capacity constraint could be binding for at least one firm, say $q_1 \leq \overline{q}_1 < D(c)$.

To go further we have to respecify demand. The reason is that if now one firm undercuts the price of its competitor, and that firm has an effective capacity constraint, then even though $p_1 < p_2,\ D_1(p_1, p_2) \neq D(p_1)$ (since $D(p_1)$ might be $> \overline{q}_1$).

What we have to do to proceed is specify who gets the low cost good when not everyone can; i.e. we need a rationing mechanism. This because the residual demand faced by the higher priced firm will depend on precisely which consumers get the lower priced good. Different rationing rules have been introduced in the literature, though the one that seems most popular is higher valuation consumers get the good first[2]. What this effectively does is "lop off" the first $\overline{q}_1$ consumers. So shift the vertical axis of the demand curve to $q = \overline{q}_1$, relabel the horizontal axis "zero" at that point, and call the demand curve which results the residual demand curve of the higher priced firm.

Now assume firm 1 plays p_1 where $D(p_1) > \overline{q}_1$ (so the first firm can not supply market demand at this price). Firm 2 has two options; it can play a price less than p_1 or play a price greater than p_1. If greater than p_1 it will play a price that maximizes against its residual demand curve (at least if it has sufficient capacity). Is this a "Nash" equilibrium? We have to check, if:

- conditional on firm 1's price, firm 2 wants to play a price below firm 1's (to do so we: give firm 2 the whole demand curve or its capacity, whichever is smallest; have the firm choose the profit maximizing price conditional on it being lower than p_1; and compute the resultant profits), and

- conditional of firm 2's price, find out if firm 1 has an incentive to play a higher price (it would never play a lower price since that would generate less profits).

Depending on the shape of the demand curve and the capacity constraint, you can get no (p_1, p_2) equilibria, a unique such equilibria, or more than one such equilibria.

Kreps and Schenkmen (1983, Bell Journal) extend the analysis into a full information two stage game; in the first stage you chose capacities, and in the second stage you chose prices conditional on capacity. They use the rationing rule above, constant marginal costs, a fixed cost per unit of capacity build, and a concave demand curve. They use a clever argument to show that the equilibrium in the two stage game is the equilibrium from the one stage Cournot game; thereby providing some support for Cournot in a world where it looks like firms set prices.

To get a bit of the intuition note that no matter the quantities chosen in period one, the Nash equilibrium price vector in period 2 is $p = D^{-1}(\overline{q}_1 + \overline{q}_2)$ provided this is greater than mc. they

[2]Another is to assume that each agent gets an equal fraction of what its demand would have been were price at the low level.

would never build capacity s.t. this would be less than mc as it would provide no benefits; i.e. they would not sell at a price less than marginal cost. If price were higher then this, firm 1 could lower its price by ϵ and receive an increment in profits of all the quantity it could provide times $\approx p$ (loosing only ϵ times its initial sales, and this can be made arbitrarily close to zero). If price were lower than this price could be raised and there with no loss in quantity. Since the Nash price is the price determined by throwing all the capacity in the market, this is essentially a quantity setting game, and one can show that a Nash equilibrium in quantity would result (i.e. we play as if the cost curve was the cost of capacity plus c and play quantities that maximize the profits from this rule).

The Kreps Schenkman assumptions are extreme. In particular a marginal cost curve which is constant up to the capacity constraint, and has infinite slope thereafter is questionable. The basic idea is that something about "scale" is determined earlier on, and though it may be changed later, the change is costly (in terms of rising marginal cost or a cost of inventory). This is true in many industries. For example in autos the firms decide which plants have stamping machines for which vehicles early on. They then determine how many shifts at the plants as information on demand roles in. For smaller changes in demand, they will use (overtime)... The adjustments are discontinuous, but respond to market conditions. Of course at the same time as they are modifying production they are modifying price, and there is movement back and forth, with "days of inventory" taking unexpected shocks. That is the actual interactions between controls and equilibrating forces is, not surprisingly, more complicated then our simple models indicate.

A Further Note on Equilibrium Notions. The Nash in price and Nash in quantity concepts have been the standard tools of static applied analysis. As noted they are not rich enough for many applied situations, particularly those where some form of dynamics comes into play. Indeed though, as we will see, our models do incredibly well in analyzing the distribution of prices in a market, we do much less well in analyzing the movements in price over time. This generates all sorts of mini literatures, like the literature on exchange rate pass through, or the literature on "sticky" prices.

When facing a particular market setting one should keep an open mind to how to model the price setting mechanism, and analyze its implications, as this is often a rich area for research. A good example of where relaxing standard assumptions can throw light on market outcomes is in a recent paper by Leslie, and Sorenson (forthcoming *Restud*) which investigates concert ticket re-sale (there is re-sale in many other markets as well). "Scalping" was illegal in many jurisdictions, at least until recently, which seems like a strange response to a mutually beneficial relationship. The worry stemmed from the fact that were re-sale allowed rent seeking activity in the primary market would diminish the welfare generated by the event and distort investment incentives. This has to be placed against the potential increase in welfare from trade that allows individuals to re-optimize given new information and the heterogeneity in value of time in going on the primary market. To evaluate the impacts of re-sale they have to compute equilibria with and without re-sale. In the equilibria with re-sale there are brokers who enter the market not because they want to view the event, but rather to purchase early and sell later at a markup. Notice that the different institutions generate different producer surplus and consumer surplus, as well as distributive effects.

We note that the institutions for re-sale of tickets entertainment events has changed recently with an increasing amount of web sites for ticket re-sale, and an increase in the use of auctions in the primary market. The implications of this are also being explored (e.g., Budish and Bhave, working paper).

2.1.3 Strategic Complements and Strategic Substitutes: The Cournot Model.

An often asked question is the following. When something changes in the environment determining the profitability of the actions of one firm, and as a result that firm, say increases, its control (in this case quantity), will other firms respond by increasing or decreasing their control? If another firm reacts by increasing its control we say the controls of the two firms are strategic complements, if that firm reacts by decreasing its control, we say the two firms controls are strategic substitutes. These terms were introduced by Bulow, Genanakopolis, and Klemperer (1985, JPE) who also discuss several applications.

Whether or not controls are strategic complements or substitutes in a given situation has important implications on the likely impacts of just about any environmental changes. For example, when there is a merger and as a result of the change in incentives generated by the change in ownership we think the merged firms will decrease their quantities, the question of the impact of the merger on price, and hence on consumer welfare will depend critically on what the other firms (i.e. the firms not in the merger) will do in response to the decrease in quantities of the merged firms. If they also decrease their quantities, the impact on price is likely to be even more adverse; if they increase their quantities they will ameliorate the impact of the merger on prices. Alternatively when there is a tariff (or a voluntary export restraint, or a tax) which effects the costs and hence the control (quantities or prices) of one competitor, the impact on prices as a whole will depend on the response of the other competitors to the increase in control of the competitor whose costs are effected.

A couple of points should be kept in mind when dealing with these concepts;

- To find the answer to the question of whether two firm's controls are strategic complements or strategic substitutes we would have to resolve say J-equations (the price equations) in J unknowns (the prices), the strategic complement/substitute concept is a *pairwise concept*. I.e. two firms in one market may have their static controls being strategic complements with respect to each other, whereas another couple of firms in the same market may have controls which are strategic substitutes.

- Also two firms can be strategic complements at one distribution of costs and set of demand conditions, and strategic substitutes at another. That is the concept is specific to a particular equilibrium configuration, and can change over configurations.

As a result whether controls are strategic complements or strategic substitutes is a matter of functional forms and the equilibrium being played; in many cases then it can not be answered without some empirical work. "Intuitive" generalizations from simpler models can easily be wrong (we come back to this when we consider Nash in prices equilibrium, where simple functional forms indicate that prices are strategic complements, but more realistic functional forms often indicate that this is not so). Things get even more complicated in dynamic models where the question is often quite important but even simple functional forms rarely generate an analytic result. E.g. Wei Tan's (2006; *Review of Industrial Organization*) analysis of the impact of the reduction in advertising forced by the Masters Settlement Agreement on cigarette demand by minors (there was a ban on advertising that was directed at minors). Paper's claim is advertising and prices are strategic complements, advertising goes down so does price, and this increases demand by minors disproportionately since they are disproportionately price sensitive.

Consider two rivals setting controls, say (x_1, x_2) (these are usually either price or quantity in static models), and assume that equilibrium is Nash. Profits of the two firms are given by

$\pi_i(x_i, x_{-i}; \cdot)$. The question of concern is whether agent i increases or decreases its x in response to an increase in the x of its rival. Assume that both before and after the change that induced the rivals price increase, all choices are interior, and satisfy a f.o.c., and that the market equilibrium is always unique.

The f.o.c. both before and after the change in the rival's control is

$$\frac{\partial \pi_i}{\partial x_i} = 0.$$

while the second order condition insures

$$\frac{\partial^2 \pi_i}{\partial x_i^2} < 0.$$

Totally differentiating the foc w.r.t. x_i we get

$$\frac{\partial^2 \pi_i(x_i, x_{-i})}{\partial x_i^2} dx_i + \frac{\partial^2 \pi_i(x_i, x_{-i})}{\partial x_i \partial x_{-i}} dx_{-i} = 0.$$

Consequently

$$\frac{dx_i}{dx_{-i}} = [\frac{\partial^2 \pi_i(x_i, x_{-i})}{\partial x_i \partial x_{-i}}]/[-\frac{\partial^2 \pi_i(x_i, x_{-i})}{\partial x_i^2}].$$

Since the denominator has to be positive for the initial x choices to be a Nash equilibrium, the sign of the l.h.s is the sign of the numerator; or the strategies are strategic substitutes at a point if the cross partial of the profit function is negative at that point, and are strategic complements if that cross partial is positive.

Strategic Complements and Strategic Substitutes;
Homogeneous Product Market with Nash in Quantities Equilibrium.

The profit functions is

$$\pi(\cdot) = p(Q)q - c(q).$$

which gives the f.o.c.

$$p(Q) - mc(q) + p'(Q)q = 0.$$

The needed cross partial is then

$$p'(Q) + p''(Q)q,$$

which, unless $p(\cdot)$ is exceptionally convex, is negative. This generates the standard intuition that in homogeneous goods models for which the appropriate equilibrium concept is Cournot, quantities are strategic substitutes. For markets which are nearly homogeneous goods markets we expect the actions of firms which are outside of the merger to partly counterbalance the effect of the merger on price.

Note that $p(\cdot)$ cannot be too convex else the original point was not an equilibrium. I.e. the second order condition for the initial choice to be an equilibrium was

$$2p'(Q) + p''(Q)q - \frac{\partial mc_j}{\partial q_j} < 0.$$

Questions.

- Assume constant marginal costs, and show that there are degrees of convexity that would generate equilibria in which the quantities are strategic complements (recall that second order conditions must be satisfied at an equilibrium).
- Assume an isoelastic demand curve and initially constant and equal marginal costs. Work out equilibrium price responses to a tariff which only effects one of the firm's marginal costs (i.e. there is a domestic and a foreign producer).

The discussion thus far has dealt with the concept of strategic complements and strategic substitutes in the context of smoothly differentiable static profit functions with unique equilibria. This concept has been extended in many ways;

- Allow for corners (so a plant can be mothballed). Here the work uses the concept of supermodularity which is an extension of "cross-partials" to situations where cross partials don't exist.
- Allow for non-uniqueness.
- Allow for dynamics.

See Milgrom and Shannon (Econometrica, 1994), and the literature cited there. The concept of strategic complements and substitutes becomes quite important in dynamics. For example, the question of whether capacity is a strategic complement or substitute is playing a big role in the debate on whether the response to entry by an incumbent airline is to increase its capacity. If this is not the "natural response", i.e. if this response is only profitable were the entrant to exit, then the DOJ would call the observation that the incumbents did increase capacity, "predatory", and there would be a case to be made. If this were the optimal response even if the entrant were to stay, say because the demand curve for the incumbent is now more elastic, then there would be no case against the incumbents. However to prove whether a pair of dynamic strategies are strategic complements or substitutes is often quite challenging.

2.2 Models of Product Differentiation

We now move away from homogeneous goods towards markets with product differentiation. This part will be much briefer.

Broadly speaking, goods can be seen as being differentiated along the following dimensions:

- Address vs. Non-Address Models:
  - Address model: Consumers prefer goods close to their "address" or location. In these types of models, an address can represent a physical location or a theoretical location in some sort of ideal space (e.g., characteristics). [E.g., local competition]
  - Non-address model: Product is a substitute for every other good in the market. [E.g., global competition]

- Vertical vs. Horizontal Models:
  - Horizontal Differentiation: Consumers differ in preferred products at the time price; i.e., at same price more than one good may be sold. Captured by having consumers differ in terms of tastes and/or preference for variety
  - (Pure) Vertical Differentiation: Consumers agree on which products are better/more preferred than others; i.e., at same price, only one good is sold.All consumers rank products' non-price attributes similarly, but different goods may be sold as consumers may differ between how they trade of "quality" with price.

Horizontal Differentiation

Example (Hotelling 1929):

- Consumers distributed uniformly along a "linear city" of length 1. Consider two goods located at the end points of the city ($x = 0$ and $x = 1$). Consumers have transportation cost t per unit of length they must travel.
- Consumer with coordinate x derives utility (if consumes) of

$$U = \begin{cases} s - p_1 - tx & \text{if buys from store 1} \\ s - p_2 - t(1 - x) & \text{if buys from store 2} \end{cases}$$

- There is an $\tilde{x}$ who is indifferent between stores:

$$\tilde{x}(p_1, p_2) = (p_2 - p_1 + t)/(2t)$$

- As long as prices are "not too high" and price difference between two shops doesn't exceed transportation cost t along city, generates demand $D_1(p_1, p_2) = N\tilde{x}$ and $D_2(p_1, p_2) = N(1-\tilde{x})$

See also (Salop 1979).

Vertical Differentiation

Example:

- Consumers buy one or zero units of a good. Goods characterized by quality index s. Utility of a consumer given by:

$$U = \begin{cases} \theta s - p & \text{if he buys a good with quality } s \text{ for price } p \\ 0 & \text{otherwise} \end{cases}$$

  θ is a taste parameter distributed according to density $f(\theta)$ with CDF $F(\theta)$ with support $[0, \infty]$.

- Two interpretations: either consumers have different tastes for quality, or they have different MRS between income and quality.

- If there is only one good, then demand for the good is simply:

$$D(p) = N[1 - F(p/s)]$$

  where N is the mass of consumers.

- With two goods, can show:
  - If $s_2/p_2 \geq s_1/p_1$ (good 2 delivers more quality per dollar than 1), only 2 will be consumed, and demand is as above (w/ one good);
  - Otherwise, let $\tilde{\theta} \equiv (p_2 - p_1)/(s_2 - s_1)$. All consumers with $\theta \geq \tilde{\theta}$ buy good 2, those with $\theta \in [p_1/s_1, \tilde{\theta})$ buy good 1, and the rest don't consume. Hence, $D_2(p_1, p_2) = N[1 - F(\tilde{\theta})]$ and $D_1(p_1, p_2) = N[F(\tilde{\theta}) - F(p_1/s_1)]$

2.3 Differentiated Products Bertrand Pricing

Just to recap: assume firms are differentiated and compete a la Bertrand in prices. The pricing rule of a monopolist is to maximize profits (assuming constant marginal costs for now):

$$\pi_{jt} = (p_{jt} - c_{jt})q_{jt}(\boldsymbol{p}) \tag{1}$$

where $\boldsymbol{p}$ is the vector of all prices. The F.O.C. for each firm of this problem is:

$$\begin{aligned}
\frac{\partial \pi_{jt}}{\partial p_{jt}} &= q_{jt}(\boldsymbol{p}) + (p_{jt} - c_{jt})\frac{\partial q_{jt}(\boldsymbol{p})}{\partial p_{jt}} \\
p_{jt} &= c_{jt} - q_{jt}\frac{\partial p_{jt}}{\partial q_{jt}} \\
p_{jt} &= c_{jt} - p_{jt}\frac{1}{\eta_{jj}} \\
(1 + 1/\eta_{jj})p_{jt} &= c_{jt}
\end{aligned}$$

Thus, if we could estimate the price elasticity, observed prices, and assumed that firms optimally set prices, we could "infer" or recover marginal costs c_{jt}.

3 Approaches to demand estimation

Approaches breakdown along the following lines:

- single vs multi-products
- within multi-product: whether you use a product space or characteristic space approach
- representative agent vs heterogenous agent
- Other breakdowns: continuous vs. discrete choice, horizontal vs. vertical, dynamic vs. static...

We will primarily focus on multi-product, demand systems with heterogeneous agents. We will cover both product and characteristics space approaches. We will focus on static settings, and later discuss methods for dealing with dynamics.

4 On Demand Estimation

4.1 Data... (briefly)

As always, the credibility and success of empirical work will hinge on the data that is leveraged. Depending on the industry and the application, data may be plentiful or sparse; it is always preferable to rely on richer data (when availabile and accessible at reasonable cost (both time and financial)) to inform our estimates than to implicitly assume them through structure or assumptions. That said, research is all about navigating these tradeoffs (and being explicit and honest about them).

To anchor disucssion, the data that we should have in mind when discussing demand estimation tends to look as follows:

- The unit of observation will be quantity of product purchased (say 12 oz Bud Light beer) together with a price for a given time period (say a week) at a location (Store, ZIP, MSA, state, country...).
- You will generally need to take a stance on the relevant market and set of products within a consumer's choice set; in addition, there typically is an outside good (e.g., non purchase) that you will need to control for (either with data or via assumptions).
- There is now a large amount of consumer-level purchase data collected by marketing firms (for instance the ERIM panel used by Ackerberg RAND 1997 to look at the effects of TV ads on yogurt purchases). However, the vast majority of demand data is aggregated at some level. As we will discuss, less-aggregated data tends to allow us to estimate more detailed (ambitious) models.
- Note that you often have a lot of information: you can get many characteristics of the good (Alcohol by volume, calories, etc) from the manufacturer or industry publications or packaging since you know the brand. The location means we can merge the demand observation with census data to get information on consumer characteristics. The date means we can look at see what the spot prices of likely inputs were at the time (say gas, electricity etc).
- Typical data sources: industry organizations, marketing and survey firms (e.g. AC Nielson), proprietary data from manufacturer, marketing departments have some scanner data online (e.g. Chicago GSB).
- The survey of consumer expenditures also has some information on person-level consumption on product groups like cars or soft-drinks.
- More often than not, data will require some ingenuity, luck, and a lot of elbow grease to obtain. Theory can help fill in some holes, but at the end of the day, good data (and variation!) is necessary for a convincing paper.

4.2 Basics: Endogeneity of Prices and Other Definitions

Consider a market equilibrium in a competitive market with the following components:

Aggregate Demand. Say it takes a constant elasticity form, i.e.

$$ln(Q_n) = x_n\beta - \alpha ln(p_n) + \epsilon_n$$

where n indexes markets, x are *observed* and ϵ are *unobserved* (by the econometrician) factors that cause differences in demand at a given price. E.g.,: parameters of income distribution, price of substitutes or complements, environmental factors that cause differences in the demand for the good,...

Aggregate supply.

$$mc_n = w_n\gamma + \lambda Q_n + \omega_n$$

w are observed and ω are unobserved (by the analyst) factors that cause differences in marginal cost. The marginal cost curve is the marginal cost of the market maker; it need not be the true social marginal cost.

Equilibrium. We assume the market is in equilibrium, i.e. demand=supply, or that the auctioneer sets price at a level where the quantity it induces equates demand and supply

$$p_n = mc_n.$$

Note that under an auctioneer interpretation, this assumes that he knows (ϵ, ω) even More generally there often are variables that are either observed to all agents, or revealed while finding the equilibrium price, that we do not contain good measures of in our data sets.

Keep in mind that:

- if there are differences in ϵ or in ω that are not known by the "auctioneers" (i.e. not incorporated in price) then there can be excess demand or supply. You can introduce that into your model, but you need a way of dealing with it. In many markets you could introduce inventories (though then you might want to add dynamics) or a rationing system. One of the important facts about electricity generation is that it is very hard (though not impossible) to store energy, and this rules out inventories. What the market maker does in electricity generation is have a special reserve market where the ISO pays a "holding" fee to generators, and can bring them up or down from a central computer to make sure the market balances at all times.
- we have simplified by assuming that last period's price does not effect either marginal cost or demand (in keeping within the simple static framework). As noted in the first lecture there are many reasons why it might, but this would put us into a world where demand or supply today depends on past, and perceptions of future, prices. I.e. a world where to analyze the determinants of current price and quantity determinants we need dynamics.

4.3 Single Product Demand Estimation

Let's now move away from competitive markets, and abstract from the supply side for a moment.

- Begin with one homogenous product. Assume demand for product j in market t could be given by $q_{jt} = D(p_{jt}, X_{j,t}, \xi_{jt})$, where q_{jt} are quantities, p_{jt} are prices, X_{jt} are exogenous variables, and ξ_{jt} are random shocks.
- Let's assume now demand is iso-elastic:

$$\ln(q_{jt}) = \alpha_j \ln p_{jt} + X_{jt}\beta + \xi_{jt} \tag{2}$$

so that price elasticity $\eta_{jt} = \alpha_j$. X_{jt} could just be an intercept for now (constant term) or a vector of demand shifters. ξ_{jt} is a one-dimensional unobserved component of demand.

Problem 1: Endogeneity of Prices

Recall from the monopoly discussion that we might be interested in price elasticities: doing so would allow us to use theory to perhaps recover ("infer") marginal cost by simply observing the price charged in a market.

- Suppose we are in a situation where the error term ξ_{jt} is correlated with higher prices (p_{jt}), i.e. $E(\xi_{jt}p_{jt}) > 0$.
- Let's decompose this correlation into:

$$\xi_{jt} = \lambda p_{jt} + \epsilon_{jt}$$

where ϵ_{jt} is the remaining uncorrelated part, and λ will typically be positive. Then we can put this back in:

$$\begin{aligned} \ln(q_{jt}) &= \alpha_j p_{jt} + X_{jt}\beta + \xi_{jt} \\ &= \alpha_j p_{jt} + X_{jt}\beta + \lambda p_{jt} + \epsilon_{jt} \\ &= \underbrace{(\alpha_j + \lambda)}_{\hat{\alpha_j}} p_{jt} + X_{jt}\beta + \epsilon_{jt} \end{aligned}$$

So the coefficient that we estimate denoted $\hat{\alpha_j}$ will be biased upwards. This will lead to unrealistically low estimates of price elasticity. We call this the *simulataneity* problem. The simultaneity (or endogeneity) problem is a recurrent theme in Empirical I.O.

- In I.O. we almost never get experimental or quasi-experimental data.
- Unlike what you've been taught in econometrics, we need to think very hard about what goes into the "unobservables" in the model (try to avoid the use of the word error term, it masks what really goes into the ϵ's in I.O. models).
- Finally, it is a *very strong* assumption to think that the firm does not react to the unobservable because it does not see it – just because I don't have the data doesn't mean a firm doesn't!
- Remember that these guys spend their lives thinking about pricing.
- Moreover, won't firms react if they see higher than expected demand yesterday?
- Note: From here on, when you are reading the papers, think hard about "is there an endogeneity problem that could be generating erroneous conclusions, and how do the authors deal with this problem?"

4.3.1 Some History.

- Henry Moore (1914)'s O.L.S. analysis of quantity on price (an attempt to estimate demand curves). Finds
  - Demand curves for agricultural products sloped down
  - Demand curves for manufacturing products sloped up.
- Working's(1927) pictures. How do we connect equilibrium dots?

It is from data such as those represented by
that we are to construct a demand curve, but e
no satisfactory fit can be obtained. A line of o
will give substantially as good a fit as will a lin
other slope.

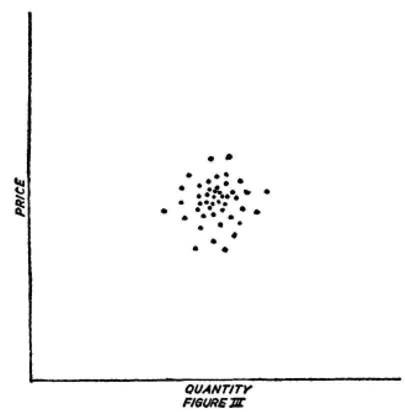

Figure 1: Working (1929 QJE)

- Needed assumption for O.L.S. on demand: $E[\epsilon|x,p] = 0$, or even $E[\epsilon(x,p)] = 0$ contradicts model and common sense (at least if the auctioneer or the firm that is pricing knows or discovers ϵ). I.e. for this to be true there is nothing that affects demand that the auctioneer knows that the empirical analyst does not know.
- Similarly needed equation for "supply" or price curve contradicts model
- Solve for price and quantity as a function of (x, w, ω, ϵ).
- Possible Solutions:
  - Estimation by 2SLS,
  - Estimation by covariance restrictions between the disturbances in the demand and supply equation.

  See any standard textbook, e.g. Goldberger(1991).

Lesson. Thought should be given to what is likely to generate the disturbances in our models, and given that knowledge we should try to think through their likely properties.

Review: What is an instrument

The broadest definition of an instrument is as follows, a variable Z such that for all possible values of Z:

$$\Pr[Z|\xi] = \Pr[Z|\xi']$$

But for certain values of X we have

$$\Pr[X|Z] \neq \Pr[X|Z']$$

So the intuition is the Z is not affected by ξ, but has some effect on X. The usual way to express these conditions is that an instrument is such that: $E[Z\xi] = 0$ and $E[XZ] \neq 0$.

Table 2
Ordinary Least Squares and Instrumental Variable Estimates of Demand Functions with Stormy Weather as an Instrument

	Ordinary least squares (dependent variable: log quantity)		*Instrumental variable*	
Variable	*(1)*	*(2)*	*(3)*	*(4)*
Log price	−0.54	−0.54	−1.08	−1.22
	(0.18)	(0.18)	(0.48)	(0.55)
Monday		0.03		−0.03
		(0.21)		(0.17)
Tuesday		−0.49		−0.53
		(0.20)		(0.18)
Wednesday		−0.54		0.58
		(0.21)		(0.20)
Thursday		0.09		0.12
		(0.20)		(0.18)
Weather on shore		−0.06		0.07
		(0.13)		(0.16)
Rain on shore		0.07		0.07
		(0.18)		(0.16)
R^2	0.08	0.23		
No. of Obs.	111	111	111	111

Source: The data used in these regressions are available by contacting the author.
Note: Standard errors are reported in parentheses.

Figure 2: Graddy (2006 JEP)

4.3.2 Representative Agent vs. Heterogeneous Agents

So far we have a representative agent model; to make it a heterogenous agent model we would have to build a micro model to make sure everything aggregated nicely, and then end up estimating something that looked something like

$$q_j = \int \gamma_i g\left(d\gamma\right) + \int \alpha_i p_j f\left(d\alpha\right) + \beta \mathbf{x}_j + \epsilon_j \tag{3}$$

Where $\alpha_i \sim F(\alpha|\theta)$ and $\gamma_i \sim G(\alpha|\tau)$ with θ and τ to be estimated. This is called a random coefficient model. Identification of the random coefficient parameters comes from differences in the sensitivity of demand to movements in price, as the price level changes. (Think about whether the model would be identified if the demand intercept were constant across all consumers). We will come back to this.

4.4 Multi-product Systems

Now let's think of a multiproduct demand system to capture the fact that most products have substitutes for each other. Generally this would be given by the relationship

$$\mathbf{q} = D(\mathbf{p}, \mathbf{X}, \xi)$$

where $\mathbf{q}, \mathbf{p}, \xi$ are $J \times 1$ vectors of quantities, prices, and random shocks, and $\mathbf{X}$ are exogenous variables. We can follow the same approach before and assume that demand takes the following isoelastic form:

$$\begin{aligned} \ln q_1 &= \sum_{j \in J} \gamma_{1j} \ln p_{1t} + \beta \mathbf{x}_{1t} + \xi_{1t} \\ &\ldots \\ \ln q_J &= \sum_{j \in J} \gamma_{Jj} \ln p_{Jt} + \beta \mathbf{x}_{Jt} + \xi_{Jt} \end{aligned}$$

4.4.1 Product vs Characteristic Space

We can think of products as being:

- a single fully integrated entity (a lexus SUV); or
- a collection of various characteristics (a 1500 hp engine, four wheels and the colour blue).

  It follows that we can model consumers as having preferences over products, or over characteristics.

The first approach embodies the product space conception of goods, while the second embodies the characteristic space approach (see Lancaster (1966, 75, 79)).

Product Space: disadvantages for estimation

[Note that disadvantages of one approach tend to correspond to the advantages of the other]

- Dimensionality: if there are J products then we have in the order of J^2 parameters to estimate to get the cross-price effects alone (the γ_{jk} terms above).
  - Can get around this to some extent by imposing more structure. For example, one can use functional form assumptions on utility: this leads to "grouping" or "nesting" approaches whereby we group products together and consider substitution across and within groups as separate things - means that ex ante assumptions need to be made that do not always make sense. More on this later.

– Can also impose symmetry: e.g., CES demand of J products with utility given by:

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

yields demand for good k:

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

where I is the income for the consumer. Note now only have to estimate ρ as opposed to number of parameters proportional to J^2. However, note this model implies:

$$\frac{\partial q_i}{\partial p_j} \frac{p_j}{q_i} = \frac{\partial q_k}{\partial p_j} \frac{p_j}{q_k} \qquad \forall i, k, j$$

which means all goods i and k have the same cross-price elasticities with respect to good j. This is an extremely strong assumption, and imposes strong restrictions on the demand system. Though popular for analytic tractability, it is not generally used in empirical IO.

- Product space methods are not well suited to handle the introduction of new goods prior to their introduction (consider how this may hinder the counterfactual exercise of working out welfare if a product had been introduced earlier - see Hausman on Cell Phones in Brookings Papers 1997 - or working out the profits to entry in successive stages of an entry game...)

Characteristic Space: disadvantages for estimation

- getting data on the relevant characteristics may be very hard and dealing with situations where many characteristics are relevant
- this leads to the need for unobserved characteristics and various computational issues in dealing with them.
- dealing with new goods when new goods have new dimensions is hard (consider the introduction of the laptop into the personal computing market)
- dealing with multiple choices and complements is a area of ongoing research, currently a limitation although work advances slowly each year.

We will explore product space approaches and then spend a fair amount of time on the characteristic space approach to demand. Most recent work in methodology has tended to use a characteristics approach and this also tends to be the more involved of the two approaches.

5 Product Space Approaches: AIDS Models

I will spend more than an average amount of time on AIDS (Almost Ideal Demand System (Deaton and Mueller 1980 AER), which wins the prize for worst acronym in all of economics models), which remain the state of the art for product space approaches. Moreover, AIDS models are still the dominant choice for applied work in things like merger analysis and can be coded up and estimated in a manner of days (rather than weeks for characteristics based approaches). Moreover, the AIDS model shows you just how far you can get with a "reduced-form" model, and these less structural models often fit the data much better than more structural models.

The main disadvantage with AIDS approaches, is that when anything changes in the model (more consumers, adding new products, imperfect availability in some markets), it is difficult to modify the AIDS approach to account for this type of problem.

- Starting point for dealing with multiple goods in product space:

$$\ln q_j = \alpha p_j + \beta \mathbf{p}_K + \gamma \mathbf{x}_j + \epsilon_j$$

- What is in the unobservable (ϵ_j)?
  - anything that shifts quantity demanded about that is not in the set of regressors
  - Think about the pricing problem of the firm ... depending on the pricing assumption and possibly the shape of the cost function (e.g. if constant cost and perfect comp, versus differentiated bertrand etc) then prices will almost certainly be endogenous. In particular, all prices will be endogenous.
  - This calls for a very demanding IV strategy, at the very least

- Also, as the number of products increases the number of parameters to be estimated will get very large, very fast: in particular, there will be J^2 price terms to estimate and J constant terms, so if there are 9 products in a market we need at least 90 periods of data!

The last point is the one to be dealt with first, then, given the specification we can think about the usual endogeniety problems. The way to reduce the dimensionality of the estimation problem is to put more structure on the choice problem being faced by consumers. This is done by thinking about specific forms of the underlying utility functions that generate empricially convenient properties. (Note that we will also use helpful functional forms in the characteristics approach, although for somewhat different reasons)

The usual empirical approach is to use a model of multi-level budgeting:

- The idea is to impose something akin to a "utility tree"
  - steps:
    1. group your products together is some sensible fashion (make sure you are happy to be grilled on the pros and cons of whatever approach you use). In Hausmann et al, the segments are Premium, Light and Standard.
    2. allocate expenditures to these groups [part of the estimation procedure].
    3. allocate expenditures within the groups [again, part of the estimation procedure]: Molson, Coors, Budweiser and etc...

Dealing with each step in reverse order:

3. When allocating expenditures within groups it is assumed that the division of expenditure within one group is independent of that within any other group. That is, the effect of a price change for a good in another group is only felt via the change in expenditures at the group level. If the expenditure on a group does not change (even if the division of expenditures within it does) then there will be no effect on goods outside that group.

	Elasticity	Standard Error
Budweiser	−4.196	0.127
Molson	−5.390	0.154
Labatts	−4.592	0.247
Miller	−4.446	0.149
Coors	−4.897	0.205
Old Milwaukee	−5.277	0.118
Genesee	−4.236	0.129
Milwaukee's Best	−6.205	0.170
Busch	−6.051	0.332
Piels	−4.117	0.469
Genesee Light	−3.763	0.072
Coors Light	−4.598	0.115
Old Milwaukee Light	−6.097	0.140
Lite	−5.039	0.141
Molson Light	−5.841	0.148

2. To be allocate expenditures across groups you have to be able to come up with a price index which can be calculated without knowing what is chosen within the group.

These two requirements lead to restrictive utility specifications, the most commonly used being the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980 AER).

5.1 Overview

This comes out of the work on aggregation of preferences in the 1970s and before. (Recall Chapter 5 of Mas-Colell, Whinston and Green)

Starting at the within-group level: assume expenditure functions for utility u and price vector p look like

$$\log(e(u,p) = (1-u)\log(a(p)) + u\log(b(p))$$

where it is assumed:

$$\begin{aligned}\log(a(p)) &= \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k\sum_j \gamma^*_{kj}\log p_k \log p_j \\ \log(b(p)) &= \log(a(p)) + \beta_0 \Pi_k p_k^{\beta_k}\end{aligned}$$

Using Shepards Lemma we can get shares of expenditure within groups as:

$$w_i = \frac{\partial log(e(u,p))}{\partial \log p_i} = \alpha_i + \sum_j \gamma_{ij}\log(p_j) + \beta_i \log\left(\frac{x}{P}\right)$$

where x is total expenditure on the group, $\gamma_{ij} = \frac{1}{2}(\gamma^*_{ij} + \gamma^*_{ji})$, P is a price index for the group and everything else should be self explanatory.

Dealing with the price index can be a pain. It can be thought of as a price index that "deflates" income. There are two ways that are used. One is the "proper" specification

$$\log(P) = \alpha_0 + \sum_k \alpha_k \log(p_k) + \frac{1}{2}\sum_j\sum_k \gamma_{kj}\log(p_k)\log(p_j)$$

which is used in the Goldberg paper, or a linear approximation (as in Stone 1954) used by most of the empirical litterature:

$$\log(P) = \sum_k w_k \log(p_k)$$

Deaton and Muellbauer go through all the micro-foundations in their AER paper.

For the allocation of expenditures across groups you just treat the groups as individual goods, with prices being the price indexes for each group. Again, note how much depends on the initial choice about how grouping works.

Steps

1. Calculate expenditure share w_i of each good i using prices p_i, quantities q_i, and total expenditure $x = \sum_k p_k q_k$.
2. Compute Stone price index: $\log P = \sum_k w_k \log(p_k)$
3. Run regression (e.g., IV):
$$w_i = \alpha_i + \sum_k \gamma_{ik} \log(p_k) + \beta_i \log(\frac{x}{P}) + \xi_i$$
where ξ_i is the error term.
4. Recover $J + 2$ parameters $(\alpha_i, \gamma_{i1}, \ldots, \gamma_{iJ}, \beta_i)$

5.2 Hausman, Leonard & Zona (1994) on Beer

This is Hausman, Leonard & Zona (1994) Competitive Analysis with Differentiated Products, Annales d'Econ. et Stat.

Here the authors want to estimate a demand system so as to be able to do merger analysis and also to discuss how you might test what model of competition best applies. The industry that they consider is the American domestic beer industry.

Note, that this is a well known paper due to the types of instruments used to control for endogeniety at the individual product level.

They use a three-stage budgeting approach: the top level captures the demand for the product, the next level the demand for the various groups and the last level the demand for individual products with the groups.

The bottom level uses the AIDS specification where spending on brand i in city n at time t is given by:
$$w_{i,n,t} = \alpha_{in} + \sum_j \gamma_{ij} \log\left(p_{jnt}\right) + \beta_i \log\left(\frac{y_{Gnt}}{P_{nt}}\right) + \varepsilon_{int}$$
where y_{Gnt} is expenditure on segment G. [note the paper makes the point that the exact form of the price index is not usually that important for the results]

The next level uses a log-log demand system
$$\log q_{mnt} = \beta_m \log y_{Bnt} + \sum_k \delta_k \log\left(\pi_{knt}\right) + \alpha_{mn} + \varepsilon_{mnt}$$
where q_{mnt} is the segment quantity purchased, y_{Bnt} is total expenditure on beer, π are segment price indices and α is a constant. [Does it make sense to switch from revenue shares at the bottom level, to quantities at the middle level?] The top level just estimates at similar equation as the middle level, but looking at the choice to buy beer overall. Again it is a log-log formulation.
$$\log u_t = \beta_0 + \beta_1 \log y_t + \beta_2 \log \Pi_t + Z_t \delta + \varepsilon_t$$

where u_t is overall spending on beer, y_t is disposable income and Π_t is a Price Index for Beer overall, and Z_t are variables controlling for demographics, monthly factors, and minimum age requirements.

Identification of price coefficients:

- recall that, as usual, price is likely to be correlated with the unobservable (nothing in the complexity that has been introduced gets us away from this problem)
- what instruments are available, especially at the individual brand level?
  - The authors propose using the prices in one city to instrument for prices in another. This works under the assumption that the pricing rule looks like:

    $$\log(p_{jnt}) = \delta_j \log(c_{jt}) + \alpha_{jn} + \omega_{jnt}$$

    where p_{jnt} is the price of good j in city n at time t, c_{jt} represents nation-wide product-costs at time t, α_{jn} are city specific shifters which reflect transportation costs or local wage differentials, and ω_{jnt} is a mean zero stochastic disturbance (e.g., local sales promotions.

    Here they are claiming that city demand shocks ω_{jnt} are uncorrelated. This allows us to use prices in other markets for the same product in the same time period as instruments (if you have a market fixed effect). Often these are referred to as *Hausman instruments*. This has been criticized for ignoring the phenomena of nation-wide ad campaigns. Still, it is a pretty cool idea and has been used in different ways in several different studies.
- Often people use factor price instruments, such as wages, the price of malt or sugar as variables that shift marginal costs (and hence prices), but don't affect the ξ's.
- You can also use instruments if there is a large price change in one period for some external reason (like a strategic shift in all the companies's pricing decisions). Then the instrument is just an indicator for the pricing shift having occurred or not.

Substitution Patterns

The AIDS model makes some assumptions about the substitution patterns between products. You can't get rid of estimating J^2 coefficients without some assumptions!

- Top level: Coors and another product (chips). If the price of Coors goes up, then the price index of beer P_B increases.
- Medium level: Coors and Old Style, two beers in separate segements. Increase in the price of Coors raises π_P, which raises the quantity of light beer sold (and hence increases the sales of Old Style in particular).
- Bottom level: Coors and Budweiser, two beers in the same segment. Increase in the price of Coors affects Budweiser through $\gamma_{c,b}$.

So the AIDS model restricts subsitution patterns to be the same between two products any two products in different segments. Is this a reasonable assumption?

TABLE 1

Beer Segment Conditional Demand Equations.

	Premium	Popular	Light
Constant	0.501	−4.021	−1.183
	(0.283)	(0.560)	(0.377)
log (Beer Exp)	0.978	0.943	1.067
	(0.011)	(0.022)	(0.015)
log ($P_{PREMIUM}$)	−2.671	2.704	0.424
	(0.123)	(0.244)	(0.166)
log ($P_{POPULAR}$)	0.510	−2.707	0.747
	(0.097)	(0.193)	(0.127)
log (P_{LIGHT})	0.701	0.518	−2.424
	(0.070)	(0.140)	(0.092)
Time	−0.001	−0.000	0.002
	(0.000)	(0.001)	(0.000)
log (# of Stores)	−0.035	0.253	−0.176
	(0.016)	(0.034)	(0.023)

Number of Observations = 101.

Figure 3: Demand Equations: Middle Level- Segment Choice

Brand Share Equations: Premium.

	1 Budweiser	2 Molson	3 Labatts	4 Miller	5 Coors
Constant	0.393	0.377	0.230	−0.104	–
	(0.062)	(0.078)	(0.056)	(0.031)	–
Time	0.001	−0.000	0.001	0.000	–
	(0.000)	(0.000)	(0.000)	(0.000)	–
log (Y/P)	−0.004	−0.011	−0.006	0.017	–
	(0.006)	(0.007)	(0.005)	(0.003)	–
log ($P_{Budweiser}$)	−0.936	0.372	0.243	0.150	–
	(0.041)	(0.231)	(0.034)	(0.018)	–
log (P_{Molson})	0.372	−0.804	0.183	0.130	–
	(0.231)	(0.031)	(0.022)	(0.012)	–
log ($P_{Labatts}$)	0.243	0.183	−0.588	0.028	–
	(0.034)	(0.022)	(0.044)	(0.019)	–
log (P_{Miller})	0.150	0.130	0.028	−0.377	–
	(0.018)	(0.012)	(0.019)	(0.017)	–
log (# of Stores)	−0.010	0.005	−0.036	0.022	–
	(0.009)	(0.012)	(0.008)	(0.005)	–
Conditional Own	−3.527	−5.049	−4.277	−4.201	−4.641
Price Elasticity	(0.113)	(0.152)	(0.245)	(0.147)	(0.203)

$$\Sigma = \begin{Bmatrix} 0.000359 & -1.436E-05 & -0.000158 & -2.402E-05 \\ - & 0.000109 & -6.246E-05 & -1.847E-05 \\ - & - & 0.005487 & -0.000392 \\ - & - & - & 0.000492 \end{Bmatrix}$$

Note: Symmetry imposed during estimation.

Figure 4: Demand Equations: Bottom-Level Brand Choice

Overall Elasticities.

	Elasticity	Standard Error
Budweiser	−4.196	0.127
Molson	−5.390	0.154
Labatts	−4.592	0.247
Miller	−4.446	0.149
Coors	−4.897	0.205
Old Milwaukee	−5.277	0.118
Genesee	−4.236	0.129
Milwaukee's Best	−6.205	0.170
Busch	−6.051	0.332
Piels	−4.117	0.469
Genesee Light	−3.763	0.072
Coors Light	−4.598	0.115
Old Milwaukee Light	−6.097	0.140
Lite	−5.039	0.141
Molson Light	−5.841	0.148

Light Segment Own and Cross Elasticities.

	Genesee Light	Coors Light	Old Milwaukee Light	Lite	Molson Light
Genesee Light	−3.763	0.464	0.397	0.254	0.201
	(0.072)	(0.060)	(0.039)	(0.043)	(0.037)
Coors Light	0.569	−4.598	0.407	0.452	0.482
	(0.085)	(0.115)	(0.058)	(0.075)	(0.061)
Old Milwaukee Light	1.233	0.956	−6.097	0.841	0.565
	(0.121)	(0.132)	(0.140)	(0.112)	(0.087)
Lite	0.509	0.737	0.587	−5.039	0.577
	(0.095)	(0.122)	(0.079)	(0.141)	(0.083)
Molson Light	0.683	1.213	0.611	0.893	−5.841
	(0.124)	(0.149)	(0.093)	(0.125)	(0.148)

Figure 5: Segment & Overall Elasticities

Merger Analysis (Preview)

Recall a single firm sets price according to

$$\frac{p_1 - mc_1}{p_1} = -\frac{1}{\eta_{11}}$$

Imagine firm owns goods $j = 1 \dots m$. Then the first order condition for the firm will be for each j:

$$\left[\frac{p_j}{\sum_{k=1}^{m} p_k q_k}\right] \frac{\partial \pi}{\partial p_j} = s_j + \sum_{k=1}^{m} \left[\frac{p_k - mc_k}{p_k} s_k\right] \eta_{kj} = 0$$

HLZ consider an hypothetical merger between two premium beers, Labatt and Coors. They find post-merger prices do not rise by too much – Coors price is constrained by Budweiser, and Labatt by Molson (another Canadian import). Without the premium beers constraining their prices, the estimates predict post-merger prices would rise by > 20%.

We will come back to these types of analysis later.

Estimated Price Increases for Hypothetical Merging Brands Assumed Efficiency Gains.

	0%	5%	10%
Coors	4.4%	−0.8%	−6.1%
	(1.2)	(1.2)	(1.1)
Labatts	3.3	−1.9	−7.0
	(1.0)	(1.0)	(0.9)

Figure 6: Merger Effects

5.3 Chaudhuri, Goldberg and Jia (2006) on Quinolones

Question: The WTO has imposed rules on patent protection (both duration and enforcement) on member countries. There is a large debate on should we allow foreign multinationals to extent their drugs patents in poor countries such as India, which would raise prices considerably.

- Increase in IP rights raises the profits of patented drug firms, giving them greater incentives to innovate and create new drugs (or formulations such as long shelf life which could be quite useful in a country like India).
- Lower consumer surplus dues to generic drugs being taken off the market.

To understand the tradeoff inherent in patent protection, we need to estimate the magnitude of these two effects. This is what CGJ do.

Market: Indian Market for antibiotics

- Foreign and Domestic, Licensed and Non-Licensed producers.
- Different types of Antibiotics, in particular CGJ look at a particular class: Quinolones.
- Different brands, packages, dosages etc...
- Question: What would prices and quantities look like if there were no unlicensed firms selling this product in the market? [3]

Data

- The Data come from a market research firm. This is often the case for demand data since the firms in this market are willing to pay large amounts of money to track how well they are doing with respect to their competitors. However, prying data from these guys when they sell it for 10 000 a month to firms in the industry involves a lot of work and emailing.
- Monthly sales data for 4 regions, by product (down to the SKU level) and prices.
- The data come from audits of pharmacies, i.e. people go to a sample of pharmacies and collect the data.
- Some products have different dosages than others. How does one construct quantity for this market?
- Some products enter and exit the sample. How can the AIDS model deal with this?

[3] One of the reasons I.O. economists use structural models is that there is often no experiment in the data, i.e. a case where some markets have this regulation and others don't.

Estimation and Results

- CGJ estimate the AIDS specification with the aggregation of different brands to product level.

  Product groups are defined to be indexed by molecule M and domestic/foreign status DF.

  Revenue share of each product group i in each region r at time t:

$$\omega_{irt} = \alpha_i + \alpha_{ir} + \sum_j \gamma_{ij} \ln p_{jrt} + \beta_i \ln(\frac{X_{Qrt}}{P_{Qrt}}) + \epsilon_{irt} \tag{4}$$

  where $\omega_{irt} = x_{irt}/X_{Qrt}$, prices for each group are aggregated/weighted over individual SKUs, and X_{Qrt} is expenditures on quinolones; and price index:

$$\ln P_Q = \alpha_0 + \sum_i \alpha_i \ln p_i + \frac{1}{2} \sum_i \sum_j \tilde{\gamma}_{ij} \ln p_i \ln p_j \tag{5}$$

  and upper level demand:

$$\omega_{Grt} = \alpha_G + \alpha_{Gr} + \sum_H \gamma_{GH} \ln P_{Hrt} + \beta_G \ln(\frac{X_{rt}}{P_{rt}}) + \epsilon_{Grt} \tag{6}$$

  across different segments H of antibiotics.

- Do not model the choice of individual SKU products:
  - Large # of SKUs within each group (dimensionality), lack of price variation at SKU level, and varying choice sets over time (entry/exit of SKUs).
  - Discrete choice approach difficult due to difficulty mapping revenue shares to physical shares – dosage of drugs not welll defined.

- Problem for the AIDS model: Over 300 different products, i.e. 90,000 cross product interaction terms to estimate! CGJ need to do some serious aggregating of products to get rid of this problem: they will aggregate products by therapeutic class into 4 of these, interacted with the nationality of the producer. I.e., each product will have an own price coefficient $\gamma_{i,i}$, and a price coefficient for products of different molecules and/or nationalities, denoted $\gamma_{i,10}, \gamma_{i,01}, \gamma_{i,00}$. (Note that these coefficients are not whether or not the molecule is licensed). Thus, a product i will exhibit the same cross-price elasticity for two different drugs if those two drugs differ in the same way both in molecule and foreign/domestic status. This yields 7 product groups (one group is only produced by foreign firms), and 7×4 price terms.

- Simultaneity bias: SKU revenue share weights (used in computation of price index for each product group) depend on expenditure, and will be correlated with demand shock. Instruments: # SKUs within group (violated if # of SKUs affect perceived quality of drug or is correlated with advertising), prices at SKU level (due to price controls)

- Supply Side: You can get upper and lower bounds on marginal costs by assuming either that firms are perfect competitors within the segment (i.e. $p = mc$) or by assuming that firms are operating a cartel which can price at the monopoly level (i.e. $p = \frac{mc}{1+1/\eta_{jj}}$). This is very smart: you just get a worse case scenario and show that even in the case with the highest possible producer profits, these profits are small compared to the loss in consumer surplus. Often it is better to bound the bias from some estimates rather than attempt to solve the problem.

- Use estimated demand system to compute the prices of domestic producers of unlicensed products that make expenditures on these products 0 (this is what "virtual prices" mean).
- Figure out what producer profits would be in the world without unlicensed firms (just $(p-c)q$ in this setup).
- Compute the change in consumer surplus (think of integrating under the demand curve).
  - Product Variety Effect
  - Expenditure Switching effect (substitution to other types of antibiotics, not quinolones); holds fixed prices of other products
  - Reduced-competition effect: firms adjust prices upwards due to removal of domestic products

Table 3—Summary Statistics for the Quinolones Subsegment: 1999–2000

	North	East	West	South
Annual quinolones expenditure per household (Rs.)	31.25	19.75	27.64	23.59
	(3.66)	(3.67)	(4.07)	(2.86)
Annual antibiotics expenditure per household (Rs.)	119.88	84.24	110.52	96.24
	(12.24)	(12.24)	(9.60)	(9.96)
No. of SKUs				
Foreign ciprofloxacin	12.38	11.29	13.08	12.46
	(1.50)	(1.90)	(1.02)	(1.06)
Foreign norfloxacin	1.83	1.71	2.00	1.58
	(0.70)	(0.75)	(0.88)	(0.83)
Foreign ofloxacin	3.04	2.96	2.96	3.00
	(0.86)	(0.86)	(0.91)	(0.88)
Domestic ciprofloxacin	106.21	97.63	103.42	105.50
	(5.99)	(4.34)	(7.22)	(4.51)
Domestic norfloxacin	38.96	34.96	36.17	39.42
	(2.71)	(2.68)	(2.51)	(3.79)
Domestic ofloxacin	18.46	16.00	17.25	17.25
	(6.80)	(6.34)	(5.86)	(6.35)
Domestic sparfloxacin	29.83	28.29	31.21	29.29
	(5.57)	(6.38)	(6.88)	(6.57)
Price per-unit API* (Rs.)				
Foreign ciprofloxacin	9.58	10.90	10.85	10.07
	(1.28)	(0.66)	(0.71)	(0.58)
Foreign norfloxacin	5.63	5.09	6.05	4.35
	(0.77)	(1.33)	(1.39)	(1.47)
Foreign ofloxacin	109.46	109.43	108.86	106.12
	(6.20)	(6.64)	(7.00)	(11.40)
Domestic ciprofloxacin	11.43	10.67	11.31	11.52
	(0.16)	(0.15)	(0.17)	(0.13)
Domestic norfloxacin	9.51	9.07	8.88	8.73
	(0.24)	(0.35)	(0.37)	(0.20)
Domestic ofloxacin	91.63	89.64	85.65	93.41
	(16.15)	(15.65)	(14.22)	(14.07)
Domestic sparfloxacin	79.72	78.49	76.88	80.28
	(9.76)	(10.14)	(11.85)	(10.37)
Annual sales (Rs. mill)				
Foreign ciprofloxacin	41.79	24.31	45.20	29.47
	(15.34)	(8.16)	(12.73)	(6.48)
Foreign norfloxacin	1.28	1.00	0.58	0.73
	(1.01)	(0.82)	(0.44)	(0.57)
Foreign ofloxacin	54.46	31.84	35.22	31.11
	(13.99)	(9.33)	(9.06)	(7.03)
Domestic ciprofloxacin	962.29	585.91	678.74	703.81
	(106.26)	(130.26)	(122.26)	(87.40)
Domestic norfloxacin	222.55	119.71	149.18	158.29
	(38.84)	(19.45)	(26.91)	(16.26)
Domestic ofloxacin	125.02	96.21	149.36	112.05
	(44.34)	(30.11)	(52.82)	(42.59)
Domestic sparfloxacin	156.17	121.75	161.30	98.11
	(31.41)	(25.76)	(46.74)	(34.20)

Note: Standard deviations in parentheses.
 * API: Active pharmaceutical ingredient.

Figure 7: Summary Statistics

Table 6A—Demand Patterns within the Quinolones Subsegment: Unconditional Price and Expenditure Elasticities in the Northern Region

Product group	Elasticity with respect to: Prices of foreign product groups: Cipro	Norflo	Oflo	Prices of domestic product groups: Cipro	Norflo	Oflo	Sparflo	Overall quinolones expenditure
Foreign ciprofloxacin	−5.57*	−0.13†	−0.15*	4.01*	0.11†	0.11†	0.16*	1.37*
	(1.79)	(0.07)	(0.07)	(1.84)	(0.06)	(0.06)	(0.06)	(0.29)
Foreign norfloxacin	−4.27†	−0.45	−4.27†	3.50†	−6.02	4.51*	4.65*	2.20*
	(2.42)	(1.12)	(2.42)	(2.10)	(6.23)	(1.84)	(1.83)	(1.05)
Foreign ofloxacin	−0.11*	−0.10†	−1.38*	−0.09	0.09†	0.23	0.11*	1.16*
	(0.05)	(0.05)	(0.31)	(0.27)	(0.05)	(0.28)	(0.04)	(0.17)
Domestic ciprofloxacin	0.18*	0.01*	−0.01	−1.68*	0.08*	0.08*	0.10*	1.17*
	(0.08)	(0.00)	(0.01)	(0.23)	(0.02)	(0.02)	(0.02)	(0.03)
Domestic norfloxacin	0.04*	−0.03	0.04*	0.58*	−2.23*	0.42*	0.40*	0.73*
	(0.01)	(0.03)	(0.01)	(0.17)	(0.11)	(0.04)	(0.03)	(0.09)
Domestic ofloxacin	0.05*	0.05*	0.11	0.77*	0.74*	−3.42*	0.74*	0.89*
	(0.02)	(0.02)	(0.13)	(0.28)	(0.08)	(0.25)	(0.08)	(0.21)
Domestic sparfloxacin	0.07*	0.04*	0.07*	1.15*	0.63*	0.63*	−2.88*	0.28*
	(0.02)	(0.01)	(0.02)	(0.15)	(0.06)	(0.06)	(0.17)	(0.12)

Notes: Standard errors in parentheses. Elasticities evaluated at average revenue shares. Asterisk (*) denotes significance at the 5-percent significance level, and dagger (†) denotes significance at the 10-percent level.

Figure 8: Elasticity Estimates

Table 7—Upper and Lower Bounds for Marginal Cost, Markup, and Annual Profit by Product Groups within the Quinolone Subsegment

Product group	Lower bound for MC (Rs.)	Upper bound for markup	Upper bound for profit (Rs. mill)	Upper bound for MC (Rs.)	Lower bound for markup	Lower bound for profit (Rs.)
Foreign ciprofloxacin	8.3*	19%	26.9	10.3	0%	0.0
	(1.23)	(0.12)	(16.55)			
Foreign norfloxacin	NA	NA	NA	5.3	0%	0.0
Foreign ofloxacin	32.3	70%*	106.1*	108.5	0%	0.0
	(23.16)	(0.21)	(31.85)			
Domestic ciprofloxacin	4.7*	59%*	1,701.9*	11.2	0%	0.0
	(1.14)	(0.10)	(298.58)			
Domestic norfloxacin	5.2*	43%*	280.7*	9.0	0%	0.0
	(0.20)	(0.02)	(15.32)			
Domestic ofloxacin	58.7*	34%*	161.2*	90.1	0%	0.0
	(2.18)	(0.02)	(12.80)			
Domestic sparfloxacin	49.5*	37%*	198.5*	78.8	0%	0.0
	(1.57)	(0.02)	(11.00)			

Notes: Standard errors in parentheses. Asterisk (*) denotes significance at the 5-percent level. Estimated lower bound for foreign norfloxacin's marginal cost is negative, since the estimated price elasticity is less than one in absolute value.

Figure 9: Marginal Costs

Table 8—Counterfactual Estimates of Consumer Welfare Losses from Product Withdrawal due to the Introduction of Pharmaceutical Patents (Rs. Bill Per Year)

		Loss of variety and:	
Counterfactual scenarios: withdrawal of one or more domestic product groups	Pure loss of variety	Cross-segment expenditure switching	Within-segment price-adjustment and cross-segment expenditure switching
Only ciprofloxacin	4.98*	4.92*	7.32*
	(0.87)	(0.89)	(1.46)
Only ofloxacin	0.08	0.08	0.23*
	(0.08)	(0.08)	(0.10)
Ciprofloxacin, ofloxacin, and norfloxacin	7.52*	7.40*	12.53*
	(1.77)	(1.80)	(4.15)
Ciprofloxacin, ofloxacin, and sparfloxacin	6.14*	6.03*	10.58*
	(1.42)	(1.45)	(3.31)
All four domestic quinolones products	11.76†	11.35†	17.81
	(6.43)	(6.34)	(12.70)

Notes: Standard errors in parentheses. Asterisk (*) denotes significance at the 5-percent significance level, and dagger (†) denotes significance at the 10-percent level.

Figure 10: Counterfactuals

5.4 Estimation in the Linear Cournot Model. [Extra Notes.]

Again we go back to the case of a cross-section of markets (indexed by n), but now we begin with a linear demand curve, which in inverse form is written as:

$$p_n = x_n\beta_x - \beta_q Q_n + \epsilon_n.$$

where, as in the perfectly competitive model, ϵ and x are unobserved and observed determinants of the level of demand,

If w are observed determinants of costs, then provided $E[\epsilon|x, w] = 0$, we can estimate the parameters of the demand equation with IV techniques.

For the supply side we now have a set of J f.o.c. for each market (these take the place of the supply curve of the competitive model). Assume linear $m.c.$ (that is, variable costs are quadratic) so that

$$mc_{n,j} = w_{n,j}\gamma + \lambda q_{n,j} + \omega_{n,j}.$$

Then if $q_{n,j} > 0$, the Cournot f.o.c. $(p - mc_j - q_j(\partial p/\partial q))$ reduce to

$$p_n - w_{n,j}\gamma - (\lambda + \beta_q)q_{n,j} - \omega_{n,j} = 0.$$

Provided $E[\omega|x, w] = 0$, estimation off of these first order conditions is fairly straightforward, though there are some details you might want to keep in mind.

Here are the standard steps in estimation. Note that you have NJ conditions of the form $E[\omega_{n,j}|x_n, w_{n,j}] = 0$.

- Write

  $$\omega_{n,j}(\theta) \equiv p_n - w_{n,j}\gamma - (\lambda + \beta_q)q_{n,j},$$

  where $\theta \equiv (\gamma, \lambda, \beta_q)$.

- Find sufficiently rich vector valued function $f(x, w)$ and form the sample moment conditions

  $$G_n(\theta) = (NJ)^{-1}\sum_{n,j}\omega_{n,j}(\theta)f(x_n, w_{n,j})$$

- Search for that value of θ that makes $\|G_n(\theta)\|$ as close as possible to zero.

Question. If $G(\theta) = EG_n(\theta)$, we usually call $G(\theta)$ the limit function, and $G(\theta_0) = 0$. What does "sufficiently rich" mean in terms of $G(\theta)$; i.e. what is the identification condition for this model? Can you derive the limit distribution of the parameters estimated in this way?

Here are some of the details you want to keep in mind when engaging in such an exercise.

- Provided we assume that no matter ω every firm produces positive outputs, the moment conditions hold at any equilibrium. So to estimate the parameters of this model we do not need "to choose" among equilibrium. On the other hand to do policy analysis we might have to.

- Note that output and price depend on the whole distribution of cost shifters. This gives us more instruments (for both demand and costs) then were available for the perfectly competitive model. It also raises the question of how to use them efficiently (Chamberlain,1986).

- If this is a market where we only observe q for a particular subset of ω (say for those whose profits would be greater than fixed costs in equilibrium), even if the original sample was a draw from a population that satisfied $E[\omega|x, w] = 0$ then generally $E[\omega|x, w, q > 0] \neq 0$ and you have a selection problem. The only way out of this is to build a model of the conditions which generate $q = 0$.

- When you have data on a cross section of markets, or a given market over time, you have to decide whether the ω's of different firms in the same market are correlated. The first thing to do is to look at market (or time) averages of residuals. If they are, you need a variance correction, and could use a more efficient estimation algorithm. This problem is also implicitly present when you are analyzing a single cross section in a given market, and one has to be careful to allow for the proper assumptions on the *realizations* of the errors.

- To gain efficiency you would generally estimate the demand equation along with the f.o.c.; you should convince yourself you know how to do this - after allowing for a covariance between the demand and cost disturbances.

The Simple Linear Cournot Model and Identification.

Note that the f.o.c. for quantity do not, per se, determine the slope of either the marginal cost or the demand function. It is only by combining the slope coefficients from the demand and f.o.c. that we can identify either.

Another way to say this, is that in this linear case the f.o.c. alone cannot tell us whether the market acts "as if" it is populated by price taking firms or by Cournot competitors. One time we interpret the slope of the "pricing equation" as being the slope of the m.c. curve, and one time we interpret as the sum of the slopes of demand and cost.[4]

On the other hand if we had variables which shifted the demand curve, but not the cost curves, we could distinguish (Bresnahan, 1982, Economic Letters). This because holding Q constant, changes in the slope of the demand curve will not effect price in a price taking regime, but will in a Nash-Cournot regime. Formally, consider an inverse demand curve with interactions between quantity and demand shifters

$$p_n = x_n \beta_x - \beta_q Q_n - Q_n x_n \beta_{q,x} + \epsilon$$

Then the f.o.c. for the Cournot model becomes

$$p_n - (\beta_q + \lambda + x_n \beta_{q,x}) \times q_{n,j} - w_{n,j}\gamma - \omega_{n,j} = 0$$

but the pricing equation under p.c. *does not change*. I.e. the f.o.c. in the p.c. world does not depend on x_n, while it does in the Cournot world. In some sense this is a "nonparametric" test of the model. I.e. we find an x_n that affects the slope of the demand curve but does not effect the slope of the cost curve, and see if it interacts with the individual quantities in the f.o.c.

[4] Of course one could ask whether the "slope" of the f.o.c. is the same as the slope of the demand curve; but you may not have data on total demand, and even if one did one might worry about that being a bit dependent on functional form assumptions, the precision of your estimators, etc. As a result a slightly different literature developed.

Here are some further things one wants to keep in mind here.

- What would you expect to effect the slope of the demand equation (in contrast to the intercept) and not effect *marginal* costs? How would you effect the distribution of income to help here? How about past advertising? The literature has discussed advertising in this context (i.e. advertising makes the demand curve for a product less elastic), but the mechanism through which this occurs is not really specified. To analyze the impact of advertising we would want to specify a micro level of how it effects individual demands and then aggregate up. This is an important topic for many reasons, and we come back to it below.

- Note that it is not clear that if we reject price-taking behavior we should immediately accept the Cournot-Nash behavior.

- All of this intuits the m.c. function from price and a behavioral assumption about how price is set. I.e. *there is no cost data.* Thus it should not be surprising that we don't get an estimate of the slope of the marginal cost function without imposing a behavioral assumption. Remember there are lots of conditions when the static optimizing behavioral choices don't make alot of sense. Also you might think of what you could do if you did have cost data, though, as noted, this is rarely the case.

The "Conduct" Parameter in the Cournot Setting.

Sometimes the literature goes further than this and looks for a "conduct" parameter. The typical discussion considers the following form of a quantity setting f.o.c.

$$p(Q) + q_j \frac{\partial p}{\partial Q} \frac{\partial Q}{\partial q_j} - mc_j = 0$$

and then interprets the term $\frac{\partial Q}{\partial q_j}$ as firm j's conjecture about the response of total industry output to increases in its own output. The conjecture is then estimated as a parameter, or as a function of the firm's market share. A simple case would be

$$p(Q) + q_j \frac{\partial p}{\partial Q}(\theta_1 + \theta_2 \frac{1}{s_j}) - mc_j = 0$$

Then:

- Under competition: $\theta_1 = \theta_2 = 0$,
- Cournot: $\theta_1 = 1$ $\theta_2 = 0$,
- Prove to yourself that under joint profit maximization (monopoly): $\theta_1 = 0$ $\theta_2 = 1$

Note: If one of these three combinations is not observed, then there is no obvious interpretation of "equilibrium" that is consistent with the estimates. Moreover before one believed a value of $\theta_2 = 1$ in a market with many firms you would want to provide evidence that one can support monopoly pricing (that there is a mechanism that insure that no firm has an incentive to deviate). This suggest that you should be wary of an equilibrium interpretation of such results (there *is no* "mode of competition" that gives you a θ_2 between zero and one). You may want to "test" zero and one, but what any other value tells you is that the equilibrium interpretation is wrong (of course you should also be wary of test statistics, since something that shows up significant with a lot of data may well be insignificant economically and vice versa).

Some Additional Notes on Estimating the Cournot model.

- Fundamentally this all relies a bit heavily on functional form (cost data are never directly used, presumably because the researchers didn't have them). You should keep in mind that to really let the data tell us what type of equilibrium is a good approximation, what we would like to do is compare price to marginal cost, and ask what behavioral model could explain the difference. There are some studies which have credibly been able to do this; a study be Genoseve and Mullin(1999,RAND) on the sugar cartel is an interesting example. But for most studies there just simply isn't enough data (especially cost data). What we do then is read about the industry, impose an equilibrium notion, insure that it is not grossly at odds with the data, and then move on to whatever we want to analyze.

- Though cost data are quite rare, as noted by Gollop and Roberts,1979, sometimes input demands are available, and they can also be used to help in estimation. That is provided factor markets are competitive, we should have

$$[p + q_j \frac{\partial p}{\partial Q}][\frac{\partial f_j(x_j)}{\partial x_{j,k}}] = w_{j,k}$$

  where $f_j(\cdot)$ is the production function for firm j, $x_{j,k}$ is the input choice, and $w_{j,k}$ is the factor price. There ought to be one first order condition for each input, and this might help both in determining the nature of equilibrium, and in recovering cost. On the other hand you will have to be in a setting where estimating product-level production functions makes sense (most times we see factor choices they are on multi-product firms or plants; see below) and make explicit assumptions on how factor choices are made (again we come back to this below).

- Often the situation is worse then the situation we have focused on (a situation in which all we have is data on outputs and prices and possibly some exogenous cost and demand shifters). Sometimes we don't have the distribution of outputs. Appelbaum(1982) and Porter(1983) both work with this situation and aggregate up the first order condition to the market level. We will show you how to do this in going over extensions to Houthackker's example below.

6 Characteristic Space Approaches to Demand Estimation

Basic approach:

- Consider products as bundles of characteristics

- Define consumer preferences over characteristics

- Let each consumer choose that bundle which maximizes their utility. We restrict the consumer to choosing only one bundle. You will see why we do this as we develop the formal model, multiple purchases are easy to incorporate conceptually but incur a big computational cost and require more detailed data than we usually have. Working on elegant ways around this problem is an open area for research.

- Since we normally have aggregate demand data we get the aggregate demand implied by the model by summing over the consumers.

6.1 Formal Treatment

- Utility of the individual:
$$U_{ij} = U(x_j, p_j, v_i; \theta)$$
for $j = \{0, 1, 2, 3, ..., J\}$.

- Good 0 is generally referred to as the *outside good.* It represents the option chosen when none of the observed goods are chosen. A maintained assumption is that the pricing of the outside good is set exogenously.

- J is the number of goods in the industry

- x_j are non-price characteristics of good j

- p_j is the price

- v_i are characteristics of the consumer i

- θ are the parameters of the model

- Note that the product characteristics do not vary over consumers, this most commonly a problem when the choice sets of consumers are different and we do not observe the differences in the choice sets.

- Consumer i chooses good j when
$$U_{ij} > U_{ik} \ \forall k \quad \text{[note that all preference relations are assumed to be strict]} \tag{7}$$

- This means that the set of consumers that choose good j is given by
$$\mathbb{S}_j(\theta) = \{v | U_{ij} > U_{ik} \ \forall k\}$$
and given a distribution over the v's, $f(v)$, we can recover the share of good j as
$$s_j(\mathbf{x}, \mathbf{p}|\theta) = \int_{\nu \in \mathbb{S}_j(\theta)} f(d\nu)$$
Obviously, if we let the market size be M then the total demand is $M \times s_j(\mathbf{x}, \mathbf{p}|\theta)$.

- This is the formal analog of the basic approach outlined above. The rest of our discussion of the characteristic space approach to demand will consider the steps involved in making this operational for the purposes of estimation.

6.1.1 Aside on utility functions

- Recall from basic micro that ordinal rankings of choices are invariant to affine transformations of the underlying utility function. More specifically, choices are invariant to multiplication of $U(\cdot)$ by a positive number and the addition of any constant.

- This means that in modelling utility we need to make some normalizations - that is we need to bolt down a zero to measure things against. Normally we do the following:

1. Normalize the mean utility of the outside good to zero.
2. Normalize the coefficient on the idiosyncratic error term to 1.

   This allows us the interpret our coefficients and do estimation.

6.2 Examples (Briefly)

Anderson, de Palma and Thisse go through many of these in very close detail. In the spring, Pierre Dubois will spend more time on variations as well.

Horizontally Differentiated vs **Vertically Differentiated** - Recall: horizontally differentiated means that, setting aside price, people disagree over which product is best. Vertically differentiated means that, price aside, everyone agrees on which good is best, they just differ in how much they value additional quality.

1. Pure Horizontal Model

   - – This is the Hotelling model (n ice-cream sellers on the beach, with consumers distributed along the beach)
     – Utility for a consumer at some point captured by ν_i is

     $$U_{ij} = \overline{u} - p_j - \theta\left(\delta_j - \nu_i\right)^2$$

     where the $\left(\delta_j - \nu_i\right)^2$ term captures a quadratic "transportation cost".
     – It is a standard workhorse for theory models exploring ideas to do with product location.

2. Pure Vertical Model

   - – Used by, Shaked and Sutton, Mussa-Rosen (monopoly pricing, slightly different), Bresnahan (demand for autos) and many others
     – Utility given by

     $$U_{ij} = \overline{u} - \nu_i p_j + \delta_j$$

     – This model is used most commonly in screening problems such a Mussa-Rosen where the problem is to set (p, q) tuples that induce high value and low value customers to self-select (2nd degree price discrimination). The model has also been used to consider product development issues, notably in computational work.

3. Logit

   - – This model assumes everyone has the same taste for quality but have different idiosyncratic taste for the product. Utility is given by

     $$U_{ij} = \delta_j + \epsilon_{ij}$$

     – $\epsilon_{ij} \overset{iid}{\sim}$ extreme value type I $\left[F\left(\epsilon\right) = e^{-e^{-\epsilon}}\right]$. This is a very helpful assumption as it allows for the aggregate shares to have an analytical form.
     I.e.:

     $$Pr(U_{ij} \geq U_{ik} \forall k) = \frac{\exp(\delta_j)}{\sum_{k=0,\ldots,J} exp(\delta_k)} \tag{8}$$

     – This ease in aggregation comes at a cost, the embedded assumption on the distribution on tastes creates more structure than we would like on the aggregate substitution matrix.

- Independence of Irrelevant Alternatives (IIA): Ratio of choice probabilities between two options j and k doesn't depend on utilities of any other product. I.e.,:

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{\delta_{ij}}}{e^{\delta_{ik}}}$$

(Red bus-Blue bus issue)
- See McFadden 1972 for details on the construction.

4. Nested Logit

- As in the AIDS Model, we need to make some "ex-ante" classification of goods into different segments, so each good $j \in S(j)$.
- $U_{ij} = V_{ij} + \epsilon_{ij}$ where goods are divided into nests, and:

$$F(\cdot) = \exp(-\sum_{s=1}^{S}(\sum_{j \in S(j)} e^{-\epsilon_{nj}/\lambda_k})^{\lambda_k})$$

$\lambda_k \in (0, 1]$ is degree of independence in unobserved components within nest k (higher means more independence).
For two different goods in different segments, the relative choice probabilities are:

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}(\sum_{j \in S_k(i)} e^{V_{nj}/\lambda_k})^{\lambda_k - 1}}{e^{V_{nm}/\lambda_l}(\sum_{j \in S_l(m)} e^{V_{nj}/\lambda_l})^{\lambda_l - 1}}$$

- The best example of using Nested-Logit for an IO application is Golberg (1995) Econometrica (in the same issue as BLP on the same industry!).
- One can classify goods into a hierarchy of nests (car or truck, foreign or domestic, nissan or toyota, camry or corrola).

5. "Generalized Extreme Value Models": Bresnahan, Trajtenberg and Stern (RAND 1997) have looked at extensions of nested logit which allow for overlapping nests: foreign or domestic computer maker in one nest and high-end or standard performance level. The advantage of this approach is that there is no nead to choose which nest comes first.

6. Ken Train (2002) discusses many different models of discrete choice. This is a great reference to get into the details of how to do these procedures. Moreover we will focus on cases where we have aggregate data, but having individual level data can help you A LOT.

7. "Ideal Type" (ADT) or "Pure Characteristic" (Berry & Pakes)

- Utility given by

$$U_{ij} = f(\nu_i, p_j) + \sum_k \sum_r g(x_{jk}, \nu_{ir}, \theta_{kr})$$

This nests the pure horizontal and pure vertical models (once you make a few function form assumptions and some normalizations.

8. BLP (1996)

- This is a parameterized version of the above case, with the logit error term tacked on. It is probably the most commonly used demand model in the empirical literature, when differentiated goods are being dealt with.

$$U_{ij} = f(\nu_i, p_j) + \sum_k \sum_r x_{jk} \nu_{ir} \theta_{kr} + \epsilon_{ij}$$

6.3 Estimation from Product Level Aggregate Data

- The data typically are shares, prices and characteristics
- That is: $\{(s_j, p_j, x_j)\}_{j=1}^{J}$
- We will start by looking at the simpler cases (the vertical model and the logit) and then move onto an examination of BLP.
- Remember that all the standard problems, like price being endogenous and wider issues of identification, will continue to be a problem here. So don't lose sight of this in all the fancy modelling!

6.3.1 Illustrative Case: Vertical Model

Note that this is what Bresnahan estimates when he looks at the possibility of collusion explaining the relative dip in auto prices in 1955.

- In the vertical model people agree on the relative quality of products, hence there is a clear ranking of products in terms of quality
- The only difference between people is that some have less willingness to pay for quality than others
- Hence (recall) utility will look like

$$U_{ij} = \overline{u} - \nu_i p_j + \delta_j$$

- To gain the shares predicted by the model we need to:
    1. Order the goods by increasing p. Note that this requires the ordering to also be increasing in δ if the goods in the sample all have non-zero share. (A good with higher p and lower δ will not be purchased by anyone.)
    2. The lowest good is the outside good (good 0) - we normalise this to zero ($\overline{u} = 0$)
    3. Choose 0 if

$$\begin{aligned} 0 &> \max_{j \geq 1} (\delta_j - \nu_i p_j) \\ \text{this implies} \quad \nu_i &> \frac{\delta_1}{p_1} \end{aligned}$$

4. Hence $\mathbb{S}_0 = \left\{ \nu \middle| \nu > \frac{\delta_1}{p_1} \right\}$. Thus if ν is distributed lognormally, $\nu = \exp(\sigma x + \mu)$ where x is distributed standard normal, then choose 0 if

$$\begin{aligned} \exp(\sigma x + \mu) &\geq \frac{\delta_1}{p_1} \\ \text{or} \quad \nu &\geq \psi_0(\theta) \end{aligned}$$

where $\psi_0(\theta) \equiv \sigma^{-1}\left[\log\left(\frac{\delta_1}{p_1}\right) - \mu\right]$, that is our model has $s_0 = F(\psi_0(\theta))$, where F is standard normal

5. Similarly, choose good 1 iff $0 < \delta_1 - \nu p_1$ and $\delta_1 - \nu p_1 \geq \delta_2 - \nu p_2$, or:

$$s_1(\theta) = F(\psi_1(\theta)) - F(\psi_0(\theta))$$

more generally

$$s_j(\theta) = F(\psi_j(\theta)) - F(\psi_{j-1}(\theta))$$

for $j = 1, ...J$.

- Question: What parameters are identified in θ? What are the sources of identification for each parameter? What are the implications for cross-price elasticities?

Estimation

To complete estimation we need to specify a data generating process. We assume we observe the choices of a random sample of size n. Each individual chooses one from a finite number of cells; Choices are mutually exclusive and exhaustive.

This suggests a multinomial distribution of outcomes

$$L_j \propto \Pi_j s_j(\theta)^{n_j}$$

Hence, choose θ to maximise the log-likelihood

$$\max_\theta \quad \sum_j n_j \log[s_j(\theta)]$$

Where n_j is the count of individuals choosing the object.

Another Example: Logit

Here the utility is

$$U_{ij} = \delta_j + \epsilon_{ij}$$

where $\epsilon_{ij} \overset{iid}{\sim}$ extreme value type II $[F(\epsilon) = e - e^{-\epsilon}]$.

This yields the closed form expressions for the share of consumers who purchase inside goods j and outside good 0:

$$\begin{aligned} s_j &= \frac{\exp[\delta_j - p_j]}{1 + \sum_{q \geq 1} \exp[\delta_q - p_q]} \\ s_0 &= \frac{1}{1 + \sum_{q \geq 1} \exp[\delta_q - p_q]} \end{aligned}$$

6.4 Identification:

Identification is the key issue, always. Here we have to get all the identification off the shares. Since $s_0 = 1 - \sum_{j \geq 1} s_j$ we have J shares to use to identify $J+2$ parameters (if we let $\theta = \{\delta_1, ..., \delta_J, \mu, \sigma\}$). (you should be able to explain this with a simple diagram) Thus hit the dimensionality problem. To solve this we need more structure. Typically we reduce the dimensionality by "projecting" product quality down onto characteristics, so that:

$$\delta_j = \sum_k \beta_k x_{kj}$$

This makes life a lot easier and we can now estimate via MLE.

An alternative approach would have been to use data from different regions or time periods which would help with this curse of dimensionality. Note that we are still in much better shape that the AIDS model since there are only $J + 2$ parameters to estimate versus $J^2 + J$ of them.

6.5 Problems with Estimates from Simple Models:

Each model has its own problems and they share one problem in common:

- Vertical Model:

  1. Cross-price elasticities are only with respect to neighbouring goods - highly constrained substitution matrix.
  2. Own-price elasticities are often not smaller for high priced goods, even though we might think this makes more sense (higher income $\rightarrow$ less price sensitivity).

- Logit Model:

  1. Own price elasticities $\eta_{jj} = \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = (-\alpha p_j (1 - s_j))$. If shares are close to 0, own price elasticities are proportional to price – higher price goods have higher elasticities.
  2. Cross-price elasticities $\eta_{jk} = \frac{\partial s_j}{\partial p_k} \frac{p_k}{s_j} = \alpha p_k s_k$. This means the cross-price elasticities for a change in product k's price is the same for all other products $j \neq k$, and is solely a function of prices and shares, but not the relative proximity of products in characteristic space. This is a bit crazy for most products (e.g., cars). This is a function of the IIA assumption.

  - Note: if you run logit, and your results do not generate these results you have bad code. This is a helpful diagnostic for programming.

- Simultaneity: No way to control for endogeniety via simultaneity. This leads to the same economically stupid results that we see in single product demand estimation that ignores endogeniety (like upward sloping demand etc).

6.6 Dealing with Simultaneity

The problem formally is that the regressors are correlated with an unobservable (we can't separate variation due to cost shocks from variationdue to demand shocks), so to deal with this we need to have an unobservable component in the model.
Let product quality be

$$\delta_j = \sum_k \beta_k x_{kj} - \alpha p_j + \xi_j$$

Where the elements of ξ are unobserved product characteristics
Estimation Strategy

1. Assume n large
2. So $s_j^o = s_j(\xi_1, ..., \xi_J|\theta)$
3. For each θ there exists a ξ such that the model shares and observed shares are equal.
4. Thus we invert the model to find ξ as a function of the parameters.
5. This allows us to construct moments to drive estimation (we are going to run everything using GMM)
   - Note: sometimes inversion is easy, sometimes it is a real pain.

Example: The Logit Model
Logit is the easiest inversion to do, since

$$\begin{aligned} \ln[s_j] - \ln[s_0] &= \delta_j = \sum_k \beta_k x_{kj} - \alpha p_j + \xi_j \\ \Rightarrow \qquad \xi_j &= \ln[s_j] - \ln[s_0] - \left(\sum_k \beta_k x_{kj} - \alpha p_j \right) \end{aligned}$$

- Note that as far as estimation goes, we now are in a linear world where we can run things in the same way as we run OLS or IV or whatever. The precise routine to run will depend, as always, on what we think are the properties of ξ.
- Further simple examples in Berry 1994

More on Estimation

- Regardless of the model we now have to choose the moment restriction we are going to use for estimation.
- This is where we can now properly deal with simultaneity in our model.
- Since consumers know ξ_j we should probably assume the firms do as well. Thus in standard pricing models you will have
$$p_j = p(x_j, \xi_j, x_{-j}, \xi_{-j})$$
- Since p is a function of the unobservable, ξ, we should not use a moment restriction which interacts p and ξ. This is the standard endogeniety problem in demand estimation.

- It implies we need some instruments.
- There is nothing special about p in this context, if $E(\xi x) \neq 0$, then we need an instruments for x as well.

Some assumptions used for identification in literature:

1. $E(\xi|x, w) = 0$ x contains the vector of characteristics other than price and w contains cost side variables. Note that they are all valid instruments for price so long as the structure of the model implies they are correlated with p_j.

   Question: how do the vertical and logit models differ in this regard?

2. Multiple markets: here assume something like

$$\xi_{jr} = \xi_j + u_{jr}$$

   and put assumptions on u_{jr}. Essentially treat the problem as a panel data problem, with the panel across region not time.

7 Generalizing Demand to allow for more Realistic Substitution Patterns: BLP

- BLP is an extension to the logit model, that allows for unobserved product characteristics and, most importantly allows for consumer heterogeneity in tastes for characteristics.
- Since it is based on a solid micro foundation it can be adapted to a variety of data types and several papers have done this in particular applications.
- The single most important contribution of BLP is showing how to do the inversion in a random-coefficient logit model, that allows the error to be popped out, and thus allowing endogeniety problems to be addressed. The next most important contribution is showing that all the machinery can produce results that make a lot of sense.
- Lastly, use the NBER working paper version - it is easier to read.

Details: The Micro Model

$$U_{ij} = \sum_k x_{jk}\beta_{ik} + \xi_j + \epsilon_{\iota j}$$

with

$$\beta_{ik} = \lambda_k + \beta_k^o \mathbf{z}_i + \beta_k^u \mathbf{v}_i$$

Definitions:

x_{jk} : observed characteristic k of product j
ξ_j : unobserved characteristics of product j
$\epsilon_{\iota j}$: the logit idiosyncratic error
λ_k : the mean impact of characteristic k
$\mathbf{z}_i$: a vector of observed individual characteristics

β_k^o : a vector of coefficients determining the impact of the elements of $\mathbf{z}_i$ on the taste for characteristic x_{jk}

$\mathbf{v}_i$: a vector of unobserved individual characteristics

β_k^u : a vector of coefficients determining the impact of the elements of $\mathbf{v}_i$ on the taste for characteristic x_{jk}

- Substituting the definition of β_{ik} into the utility function you get

$$U_{ij} = \sum_k x_{jk}\lambda_k + \sum_k x_{jk}\beta_k^o \mathbf{z}_i + \sum_k x_{jk}\beta_k^u \mathbf{v}_i + \xi_j + \epsilon_{\iota j}$$

  or, as is usually the way this is written (and also the way you end up thinking about things when you code up the resulting estimator)

$$U_{ij} = \delta_j + \sum_k x_{jk}\beta_k^o \mathbf{z}_i + \sum_k x_{jk}\beta_k^u \mathbf{v}_i + \epsilon_{\iota j}$$

  where

$$\delta_j = \sum_k x_{jk}\lambda_k + \xi_j$$

- Note that this model has two different types of interactions between consumer characteristics and product characteristics:

1. (a) i. Interactions between observed consumer characteristics $\mathbf{z}_i$ and product characteristics x_{jk}'s; and

   ii. Interactions between unobserved consumer characteristics $\mathbf{v}_i$ and product characteristics x_{jk}'s

- These interactions are the key things in terms of why this model is different and preferred to the logit model. These interactions kill the IIA problem and mean that the aggregate substitution patterns are now far more reasonable (which is to say they are not constrained to have the logit form).

  - Question: Are the substitution patterns at the individual level any different from the logit model?

The intuition for why things are better now runs as follows:

- If the price of product j (say a BMW 7 series) increases, very specific customers will leave the car - those customers who have a preference for the car's characteristics and consequently will like cars close to it in the characteristic space that the empirical researcher is using.

- Thus they will substitute to cars that are close to the BMW in characteristic space (say a Lexus, and not a Reliant Regal (a three wheeled engineering horror story still sometimes seen in the UK)

- Also, price effects will be different for different products. Products with high prices, but low shares, will be bought by people who don't respond much to price and so they will likely have higher markup than a cheap product with the same share.

Figure 11: The Reliant Regal

- This model also means that products can be either strategic complements or substitutes in the pricing game. (in Logit they are strategic complements).
- Usually, we only have product level data at the aggregate level so the source of consumer information is the distribution of $\mathbf{z}_i$ from the census. That is, we are usually working with the $\mathbf{v}_i$ part of the model. However, a few studies have used micro data of one form or another, notably MicroBLP (JPE 2004).
- With micro data you need to think about whether the individual specific data you have is enough to capture the richness of choices. If not,then you need to also include the unobserved part of the model as well.

7.1 Estimation: Step by step overview

We consider product level data (so there are no observed consumer characteristics). Thus we only have to deal with the $\mathbf{v}$'s.

$$U_{ij} = \underbrace{\delta_j}_{\sum_k x_{jk}\lambda_k + \xi_j} + \sum_k x_{jk}\beta_k^u \mathbf{v}_i + \epsilon_{\imath j}$$

Step 0: Search over $\boldsymbol{\theta}$ Generally, we are looking for $\boldsymbol{\theta} \equiv \{\boldsymbol{\lambda}, \boldsymbol{\beta}\}$ that minimizes a GMM Objective (see section notes). If the distribution of ν_i is unknown but can be parameterized, these parameters are also estimated within $\boldsymbol{\theta}$.

For a given evaluation of the objective function and a given $\boldsymbol{\theta}$, we will be looking for the implied set of product "mean-utilities" $\{\delta_j\}_{\forall j}$so that the predicted shares of the model match those observed in the data. Once these are recovered, we can recover $\boldsymbol{\xi}(\boldsymbol{\theta})$ that, along with our candidate parameter vector (θ), induces our model to match the shares in the data, and then compute the GMM objective $G(\boldsymbol{\xi}(\boldsymbol{\theta}), \boldsymbol{Z}; \boldsymbol{\theta})$ where $\boldsymbol{Z}$ represent instruments.

Thus, one performs the following steps:

Step 1: Work out the aggregate shares conditional on (δ, β)

- After integrating out the $\epsilon_{\iota j}$ (recall that these are familiar logit errors) the equation for the share is

$$s_j(\delta, \beta) = \int \frac{\exp\left[\delta_j + \sum_k x_{jk}\mathbf{v}_i\beta_k\right]}{1 + \sum_{q \geq 1} \exp\left[\delta_q + \sum_k x_{qk}\mathbf{v}_i\beta_k\right]} f(\mathbf{v})\, d\mathbf{v}$$

- This integral is not able to be solved analytically. (compare to the logit case). However, for the purposes of estimation we can handle this via simulation methods. That is, we can evaluate the integral use computational methods to implement an estimator...

- Take ns simulation draws from $f(\mathbf{v})$. This gives you the simulated analog

$$\widehat{s}_j^{ns}(\delta, \beta) = \sum_r \frac{\exp\left[\delta_j + \sum_k x_{jk}\mathbf{v}_{ir}\beta_k\right]}{1 + \sum_{q \geq 1} \exp\left[\delta_q + \sum_k x_{qk}\mathbf{v}_{ir}\beta_k\right]}$$

Note the following points:

- The logit error is very useful as it allows use to gain some precision in simulation at low cost.

- If the distribution of a characteristic is known from Census data then we can draw from that distribution (BLP fits a Lognormal to census income data and draws from that)

- By using simulation you introduce a new source of error into the estimation routine (which goes away if you have "enough" simulations draws...). Working out what is enough is able to be evaluated (see BLP). The moments that you construct from the simulation will account for the simulation error without doing special tricks so this is mainly a point for interpreting standard errors.

- The are lots of ways to use simulation to evaluate integrals, some of them are quite involved. Depending on the computational demands of your problem it could be worth investing some time in learning some of these methods. (Ken Judd has a book in computation methods in economics that is a good starting point, Ali can also talk to you about using an extension of Halton Draws, called the Latin Cube to perform this task)

Step 2: Recover the ξ from the shares.

Remember from basic econometrics that when we want to estimate using GMM we want to exploit the orthogonality conditions that we impose on the data. To do this we need to be able to compute the unobservable, so as to evaluate the sample moments. So how to do this? This is one of the main contributions of BLP:

- BLP point out that iterating on the system

$$\delta_j^k(\beta) = \delta_j^{k-1}(\beta) + \ln\left[s_j^o\right] - \ln\left[\widehat{s}_j^{ns}\left(\delta^{k-1}, \beta\right)\right]$$

has a unique solution (the system is a contraction mapping with modulus less than one and so has a fixed point to which it converges monotonically at a geometric rate). Both Nevo or BLP also exploit the fact that the following is also a contraction

$$\exp\left[\delta_j^k(\beta)\right] = \exp\left[\delta_j^{k-1}(\beta)\right] \frac{s_j^o}{\widehat{s}_j^{ns}(\delta^{k-1}, \beta)}$$

This is what people actually use in the programming of the estimator.

- So given we have $\delta(\beta, s^o, P^{ns})$ we have an analytical form for λ and ξ (which we be determined by the exact indentifying assumptions you are using). In other words

$$\xi(\beta, s^o, P^{ns}) = \delta(\beta, s^o, P^{ns}) - \sum_k \mathbf{x}_k \lambda_k$$

  and depending on the moments you are using, you can "concentrate out" λ_k as a function of the non-linear parameters (see Nevo JEMS RA guide appendix; similar to OLS but need to account for potentially different GMM weighting matrix).

- The implication is that you should only be doing a nonlinear search over the elements of β.

Step 3: Construct the Moments

We want to interact $\xi(\beta, s^o, P^{ns})$ with the instruments which will be the exogenous elements of x and our instrumental variables w (recall that we will be instrumenting for price etc).

You need to make sure that you have enough moment restrictions to identify the parameters of interest.

Step $4 \to 0$: Iterate until have reached a minimum

- Recall that we want to estimate (λ, β). Given the β the λ have analytic form, we only need to search over the β that minimize our objective function for minimizing the moment restrictions.

- Look back at the expression for the share and you will realize that the β is the only thing in there that we need to determine to do the rest of these steps. However, since it enters nonlinearly we need to do a nonlinear search to recover the values of β that minimize our objective function over the moments restrictions.

- You will need to decide on a stopping point.

- Some things to note about this:

  - This means that estimation is computationally intensive.
  - You will need to use Matlab, Fortran, C, Gauss, R etc to code it up. I like Matlab, personally.
  - There are different ways to numerically search for minimums: the advantage of a simplex search algorithm over derivative based methods is that they are a bit more robust to poorly behaved functions, but take longer. Also start you code from several different places before believing a given set of results. [in matlab fminsearch is a good tool]. Also newer search methods (e.g., KNITRO) with some cost (learning, financial) have been reported to be better than built in Matlab search functions.
  - Alternatively, and even better thing to do is to use a program that can search for global minima so that you don't have to worry too much about starting values. These can take about 10-20 times longer, but at least you can trust your estimates. Some are Differential Evolution and Simulated Annealing. You can get these in MATLAB, C or FORTRAN off the web.
  - Aviv Nevo has sample code posted on the web and this is a very useful place to look to see the various tricks in programming this estimator.

- Due to the non-linear nature of the estimator, the computation of the standard errors can be a arduous process, particularly if the data structure is complex.
- Taking numerical derivatives will often help you out in the computation of standard errors.
- For more details on how to construct simulation estimators and the standard errors for nonlinear estimators, look to you econometrics classes (and the GMM section notes)

7.2 Identification in these models

The sources of identification in the standard set up are going be:

1. differences in choice sets across time or markets (i.e. changes in characteristics like price, and the other x's)
2. differences in underlying preferences (and hence choices) over time or across markets
3. observed differences in the distribution of consumer characteristics (like income) across markets
4. the functional form will play a role (although this is common to any model, and it is not overly strong here)

- so if you are especially interesting in recovering the entire distribution of preferences from aggregate data you may be able to do it with sufficiently rich data, but it will likely be tough without some additional information or structure.
- additional sources of help can be:
  - adding a pricing equation (this is what BLP does)
  - add data, like micro data on consumer characteristics, impose additional moments from other data sources to help identify effects of consumer characteristics (see Petrin on the introduction of the minivan), survey data on who purchases what (MicroBLP).

7.3 Adding in "Supply Side" Moments

One can also lean on a theoretical model of price competition in order to restrict the behavior of the estimated demand system. This is what BLP also do (recall ideas from Bresnahan 87).

A firm maximizes its profits over the set of products it produces:

$$\Pi_F(j) = M \sum_{j \in F(j)} s_{jt}(p_{jt} - mc_{jt}) \tag{9}$$

where M is market size. Taking the first-order condition (and dropping out M) you get:

$$\frac{\partial \Pi}{\partial p_{jt}} = s_{jt} + \sum_{k \in F(j)} \frac{s_{kt}}{p_{jt}}(p_{kt} - mc_{kt}) = 0 \; \forall j \tag{10}$$

Define the ownership matrix as Ω where $\Omega_{jk} = 1$(product j and k are owned by the same firm). Then we can stack all the FOCs accross all products j in market t to get:

$$\mathbf{s} + \boldsymbol{\Omega} \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}}(\mathbf{p} - \mathbf{mc}) = 0 \tag{11}$$

where $\cdot *$ is the element-by-element matrix product. Rearranging we get marginal costs:

$$\mathbf{mc} = p + (\boldsymbol{\Omega} \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}})^{-1}\mathbf{s} \tag{12}$$

We can use the supply side as an extra moment condition when estimating demand. Suppose that marginal cost as determined by:

$$\ln(mc_{jt}) = X_{jt}\gamma + \omega_{jt} \tag{13}$$

where the X's are things like car weight, horsepower and other factors that can change marginal costs. In the soft drink industry I know that all coke brands in the same bottle size have the same marginal costs, and I can impose this by having a coke brand dummy in the X's.

Since the RHS of (12) is a function of $\boldsymbol{\theta}$, we can recover an $\boldsymbol{\omega}(\boldsymbol{\theta})$ during our estimation routine (once we add $\boldsymbol{\gamma}$ to the parameters being estimated), and we can add $E(\omega(\boldsymbol{\theta})Z) = 0$ to the previous moment conditions.

Idea: As in Bresnahan 87:

7.4 Overview of BLP Results

BLP estimates this system for the US car market using data on essentially all car makes from 1971-1990. The characteristics are:

- cylinders
- # doors
- weight
- engine displacement
- horsepower
- length
- width
- wheelbase
- EPA miles per gallon
- dummies for automatic, front wheel drive, power steering and air conditioning as standard features.
- price (which is the list price) all in 1983 dollars

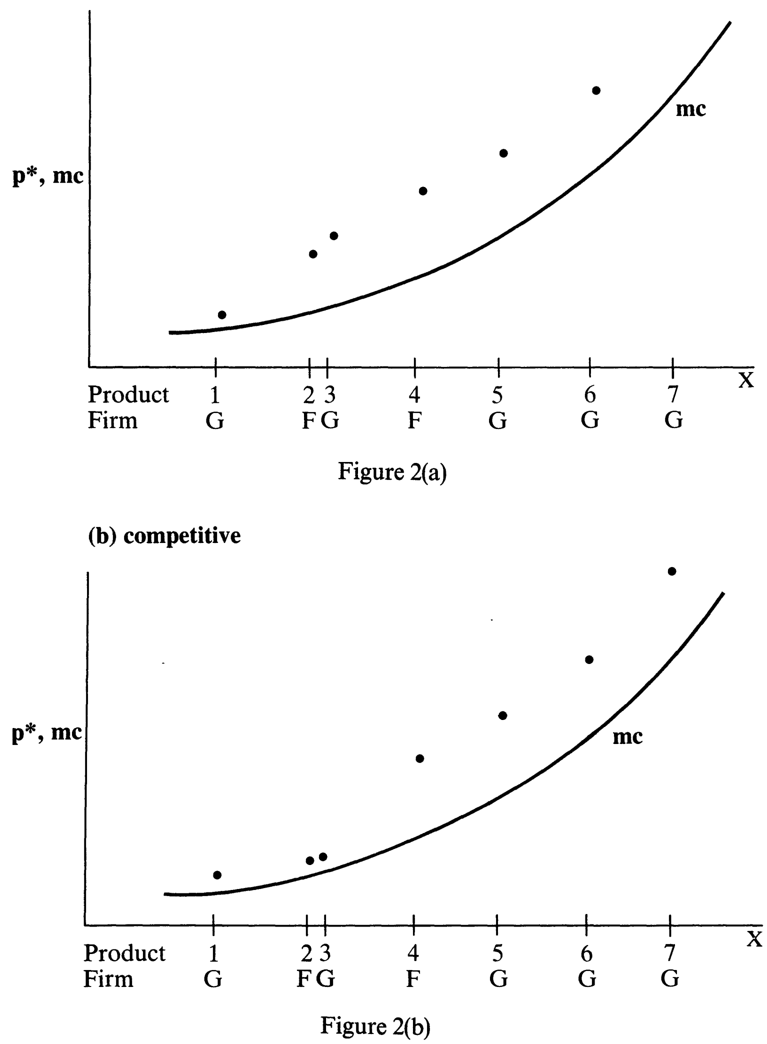

Figure 12: Bresnahan: Intuition for Supply Side Moments and Conduct.

year/model is an observation = 2217 obs
Instruments:

- Products that face substitutes will tend to have low markups, and those with poor substitutes will tend to have high markups.
- Hence, BLP motivate use of characteristics of products produced by *rival* firms and those of other products within the *same* firm as instruments.

7.5 Nevo 1998

- Nevo is trying to understand why there are such high markups in the Ready-to-Eat Cereal Industry, and where does market power come from in this industry.

TABLE IV

ESTIMATED PARAMETERS OF THE DEMAND AND PRICING EQUATIONS: BLP SPECIFICATION, 2217 OBSERVATIONS

Demand Side Parameters	Variable	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Means ($\bar{\beta}$'s)	*Constant*	−7.061	0.941	−7.304	0.746
	HP/Weight	2.883	2.019	2.185	0.896
	Air	1.521	0.891	0.579	0.632
	MP$	−0.122	0.320	−0.049	0.164
	Size	3.460	0.610	2.604	0.285
Std. Deviations (σ_β's)	*Constant*	3.612	1.485	2.009	1.017
	HP/Weight	4.628	1.885	1.586	1.186
	Air	1.818	1.695	1.215	1.149
	MP$	1.050	0.272	0.670	0.168
	Size	2.056	0.585	1.510	0.297
Term on Price (α)	$\ln(y-p)$	43.501	6.427	23.710	4.079
Cost Side Parameters					
	Constant	0.952	0.194	0.726	0.285
	ln (*HP/Weight*)	0.477	0.056	0.313	0.071
	Air	0.619	0.038	0.290	0.052
	ln (*MPG*)	−0.415	0.055	0.293	0.091
	ln (*Size*)	−0.046	0.081	1.499	0.139
	Trend	0.019	0.002	0.026	0.004
	$\ln(q)$			−0.387	0.029

Figure 13: BLP Model Estimates.

- Part of the story comes from the production side: there are only 13 plants in the U.S. that manufacture RTE cereal. Nevo will focus on the demand side.
- Data: Regional Sales of Brand name RTE cereal over several years. (Actually a fair bit of data)
- Unlike BLP, Nevo does not need to use supply side moments and uses brand dummies.

The supply side is:
A firm maximizes it's profits over the set of products it produces:

$$\Pi_F(j) = M \sum_{j \in F(j)} s_{jt}(p_{jt} - mc_{jt}) \tag{14}$$

where M is market size. Taking the first-order condition (and dropping out M) you get:

$$\frac{\partial \Pi}{\partial p_{jt}} = s_{jt} + \sum_{k \in F(j)} \frac{s_{kt}}{p_{kt}}(p_{jt} - mc_{jt}) = 0 \tag{15}$$

TABLE VI

A SAMPLE FROM 1990 OF ESTIMATED OWN- AND CROSS-PRICE SEMI-ELASTICITIES: BASED ON TABLE IV (CRTS) ESTIMATES

	Mazda 323	Nissan Sentra	Ford Escort	Chevy Cavalier	Honda Accord	Ford Taurus	Buick Century	Nissan Maxima	Acura Legend	Lincoln Town Car	Cadillac Seville	Lexus LS400	BMW 735i
323	−125.933	1.518	8.954	9.680	2.185	0.852	0.485	0.056	0.009	0.012	0.002	0.002	0.000
Sentra	0.705	−115.319	8.024	8.435	2.473	0.909	0.516	0.093	0.015	0.019	0.003	0.003	0.000
Escort	0.713	1.375	−106.497	7.570	2.298	0.708	0.445	0.082	0.015	0.015	0.003	0.003	0.000
Cavalier	0.754	1.414	7.406	−110.972	2.291	1.083	0.646	0.087	0.015	0.023	0.004	0.003	0.000
Accord	0.120	0.293	1.590	1.621	−51.637	1.532	0.463	0.310	0.095	0.169	0.034	0.030	0.005
Taurus	0.063	0.144	0.653	1.020	2.041	−43.634	0.335	0.245	0.091	0.291	0.045	0.024	0.006
Century	0.099	0.228	1.146	1.700	1.722	0.937	−66.635	0.773	0.152	0.278	0.039	0.029	0.005
Maxima	0.013	0.046	0.236	0.256	1.293	0.768	0.866	−35.378	0.271	0.579	0.116	0.115	0.020
Legend	0.004	0.014	0.083	0.084	0.736	0.532	0.318	0.506	−21.820	0.775	0.183	0.210	0.043
TownCar	0.002	0.006	0.029	0.046	0.475	0.614	0.210	0.389	0.280	−20.175	0.226	0.168	0.048
Seville	0.001	0.005	0.026	0.035	0.425	0.420	0.131	0.351	0.296	1.011	−16.313	0.263	0.068
LS400	0.001	0.003	0.018	0.019	0.302	0.185	0.079	0.280	0.274	0.606	0.212	−11.199	0.086
735*i*	0.000	0.002	0.009	0.012	0.203	0.176	0.050	0.190	0.223	0.685	0.215	0.336	−9.376

Note: Cell entries i, j, where i indexes row and j column, give the percentage change in market share of i with a $1000 change in the price of j.

Figure 14: Elasticities from the BLP Model.

TABLE VII

SUBSTITUTION TO THE OUTSIDE GOOD

Model	Given a price increase, the percentage who substitute to the outside good (as a percentage of all who substitute away.) Logit	BLP
Mazda 323	90.870	27.123
Nissan Sentra	90.843	26.133
Ford Escort	90.592	27.996
Chevy Cavalier	90.585	26.389
Honda Accord	90.458	21.839
Ford Taurus	90.566	25.214
Buick Century	90.777	25.402
Nissan Maxima	90.790	21.738
Acura Legend	90.838	20.786
Lincoln Town Car	90.739	20.309
Cadillac Seville	90.860	16.734
Lexus LS400	90.851	10.090
BMW 735*i*	90.883	10.101

Figure 15: BLP: Comparison b/w RC and Logit.

Define the ownership matrix as Ω where $\Omega_{jk} = 1$(product j and k are owned by the same firm). Then we can stack all the FOCs accross all products j in market t to get:

$$\mathbf{s} + \mathbf{\Omega} \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}}(\mathbf{p} - \mathbf{c}) = 0 \tag{16}$$

where $\cdot*$ is the element-by-element matrix product. Rearranging we get marginal costs:

$$\mathbf{c} = p + (\mathbf{\Omega}\frac{\partial \mathbf{s}}{\partial \mathbf{p}})^{-1}\mathbf{s} \tag{17}$$

We can use the supply side as an extra moment condition when estimating demand. Suppose that marginal cost as determined by:

$$\ln(mc_{jt}) = X_{jt}\gamma + \omega_{jt} \tag{18}$$

2 S. BERRY, J. LEVINSOHN, AND A. PAKES

TABLE VIII

A SAMPLE FROM 1990 OF ESTIMATED PRICE-MARGINAL COST MARKUPS AND VARIABLE PROFITS: BASED ON TABLE 6 (CRTS) ESTIMATES

	Price	Markup Over MC $(p - MC)$	Variable Profits (in $'000's) $q*(p - MC)$
Mazda 323	$5,049	$ 801	$18,407
Nissan Sentra	$5,661	$ 880	$43,554
Ford Escort	$5,663	$1,077	$311,068
Chevy Cavalier	$5,797	$1,302	$384,263
Honda Accord	$9,292	$1,992	$830,842
Ford Taurus	$9,671	$2,577	$807,212
Buick Century	$10,138	$2,420	$271,446
Nissan Maxima	$13,695	$2,881	$288,291
Acura Legend	$18,944	$4,671	$250,695
Lincoln Town Car	$21,412	$5,596	$832,082
Cadillac Seville	$24,353	$7,500	$249,195
Lexus LS400	$27,544	$9,030	$371,123
BMW 735*i*	$37,490	$10,975	$114,802

Figure 16: BLP: Markups.

where the X's are things like car weight, horsepower and other factors that can change marginal costs. In the soft drink industry I know that all coke brands in the same bottle size have the same marginal costs, and I can impose this by having a coke brand dummy in the X's.

The additional moment condition become $E(\omega Z) = 0$ which we can just add to the previous moment conditions.

8 Some Applications of Characteristics Based Demand Systems

8.1 "Micro"-BLP (2004 JPE)

Uses CAMIP data provided by GM (not available outside company). Survey is a sample of 1993 vehicle registrations, where a given # of purchasers of each vehicle are sampled. Almost all vehicles sold in US (not just GM); 37K observations. HH attributes (income, age of HoH, family size, place of residence (urban, rural, etc.),...) Tends to be higher-income than CPS.

Key in this paper is the use of "second-choice" data conditional on purchase. "Direct, data-based measure of substitution" that provides identifying power w/o exogenous changes in choice sets.

$$u_{ij} = \sum_k x_{jk}\beta_{ik} + \xi_j + \epsilon_{ij}$$

where

$$\beta_{ik} = \beta_k + \sum_r z_{ir}\beta^o_{kj} + \beta^u_k \nu_{ik}$$

TABLE III

DETAILED ESTIMATES OF PRODUCTION COSTS

Item	$/lb	% of Mfr Price	% of Retail Price
Manufacturer Price	2.40	100.0	80.0
Manufacturing Cost:	1.02	42.5	34.0
Grain	0.16	6.7	5.3
Other Ingredients	0.20	8.3	6.7
Packaging	0.28	11.7	9.3
Labor	0.15	6.3	5.0
Manufacturing Costs (net of capital costs)[a]	0.23	9.6	7.6
Gross Margin		57.5	46.0
Marketing Expenses:	0.90	37.5	30.0
Advertising	0.31	13.0	10.3
Consumer Promo (mfr coupons)	0.35	14.5	11.7
Trade Promo (retail in-store)	0.24	10.0	8.0
Operating Profits	0.48	20.0	16.0

[a] Capital costs were computed from ASM data.

Source: Cotterill (1996) reporting from estimates in CS First Boston Reports "Kellogg Company," New York, October 25, 1994.

Figure 17: Costs and Back of the envelope markups in the RTE Cereal Industry.

where $\boldsymbol{z}_i$ and $\boldsymbol{\nu}_i$ are observed and unobserved consumer attributes. Can be re-expressed:

$$u_{ij} = \underbrace{\sum_k x_{jk}\beta_k + \xi_j}_{\delta_j} + \sum_r z_{ir}\beta^o_{kj} + \beta^u_k \nu_{ik} + \epsilon_{ij}$$

With parametric assumptions on $\boldsymbol{z}$ and $\boldsymbol{\nu}$ and standard regularity conditions, can obtain consistent estimates of $\theta \equiv \{\delta, \beta^o, \beta^u\}$ from microlevel data; however, things such as price elasticities (which depend on knowing how δ changes when price changes) will require additional assumptions on ξ's. Can assume (as in BLP) that ξ is mean independent of nonprice characteristics of all the products.

Use as moments:

1. the covariance of the observed first-choice product characteristics (x) with observed consumer attributes (z) (e..g, family size and first-choice vehicle size);

2. the covariance between the first-choice product characteristics and the second-choice characteristics (e.g., covariance of the size of the first-choice vehicle with the size of the second-choice vehicle);

3. the market shares of the J products.

TABLE VI

RESULTS FROM THE FULL MODEL[a]

Variable	Means (β's)	Standard Deviations (σ's)	Interactions with Demographic Variables: Income	Income Sq	Age	Child
Price	−27.198	2.453	315.894	−18.200	—	7.634
	(5.248)	(2.978)	(110.385)	(5.914)		(2.238)
Advertising	0.020	—	—	—	—	—
	(0.005)					
Constant	−3.592[b]	0.330	5.482	—	0.204	—
	(0.138)	(0.609)	(1.504)		(0.341)	
Cal from Fat	1.146[b]	1.624	—	—	—	—
	(0.128)	(2.809)				
Sugar	5.742[b]	1.661	−24.931	—	5.105	—
	(0.581)	(5.866)	(9.167)		(3.418)	
Mushy	−0.565[b]	0.244	1.265	—	0.809	—
	(0.052)	(0.623)	(0.737)		(0.385)	
Fiber	1.627[b]	0.195	—	—	—	−0.110
	(0.263)	(3.541)				(0.0513)
All-family	0.781[b]	0.1330	—	—	—	
	(0.075)	(1.365)				
Kids	1.021[b]	2.031	—	—	—	
	(0.168)	(0.448)				
Adults	1.972[b]	0.247	—	—	—	
	(0.186)	(1.636)				
GMM Objective (degrees of freedom)			5.05 (8)			
MD χ^2			3472.3			
% of Price Coefficients > 0			0.7			

[a] Based on 27,862 observations. Except where noted, parameters are GMM estimates. All regressions include brand and time dummy variables. Asymptotically robust standard errors are given in parentheses.

[b] Estimates from a minimum-distance procedure.

Figure 18: Estimated BLP Model.

First "help" identify β^o; second with β^u. Given $\{\beta^o, \beta^u\}$, there is a unique δ that matches observed market shares to predicted shares; hence, the use of 3 to id δ.

Once δ is obtained, can try to get β_k:

$$\delta_j = p_j\beta_p + \sum_{k\neq p} x_{jk}\beta_k + \xi_j$$

Issue is that as opposed to BLP (20 annual cross sections), here only have a single 1993 cross section.

1. Could estimate w/ IV. Estimate: 2.5 w/ standard error 25.

2. Could mimic BLP and use supply side moment restrictions. Reliance on functional form and pricing assumption. Estimate of -3.58 w/ se of .22.

3. Use GM idea that aggregate (market) elasticity of new products is -1; gets $\beta_p = 11$.

Though levels of price elasticities vary, the patterns are the same: semi-elasticities decrease w/ price, and (hodling fixed price) the elasticities of pickups, vans, SUVs, and (to a lesser extent) sports cars are lower (explaining larger markups).

Uses estimates to simulate introduction of Toyota and Mercedes SUV (sets new ξ to average of that firm's products), and the shutting down of Oldsmobile division by GM.

8.2 Petrin (2002 JPE)

- We often want to quantify the benefits of innovation. Theory is often ambiguous (e.g., new product introductions to gain market power, but new products may fill unserved needs).
- You need a demand curve to do this since we want to know what people would be willing to pay above the market price.
- Brief History: Chrysler/Dodge introduced Caravan in 1984, sold 170K in first year; others (GM, Ford) tried to imitate, but didn't do as well since they relied on truck (as opposed to car) platform. Chrysler maintained large market share (44% after 14 years), but canibalized station wagon sales.
- Petrin quantifies the social benifit of the minivan: Increases total welfare by about 2.9 billion dollars from 1984-88, most of which is consumer surplus not profits which are captured by firms.

Also uses "Micro-Moments": i.e., moment conditions coming from micro-data. So for example, one might have data coming from the CEX on the average amount of money spent on soft drinks by people who earn less than \$ 10 000 a year, which I call $\hat{s}_{t|I<10000}$. The model's prediction is:

$$s_{t|I<10000,\theta} = \sum_{j>0} \frac{1}{1(I_{k<10000})} \sum_k (s_{ijt}(\theta) 1(I_{k<10000})) \tag{19}$$

So we can build an error into the model which is $\zeta_t = \hat{s}_{t|I<10000} - s_{t|I<10000,\theta}$ and treat it like all our other moment conditions.

Petrin's estimating utility eq:

$$u_{ij} = \alpha_i \ln(y_i - p_j) + X_j\beta + \sum_k \gamma_{ik} x_{jk} + \xi_j + \epsilon_{ij} \tag{20}$$

where $\gamma_{ik} = \gamma_k \nu_{ik}$ and ν_{ik} is an idiosyncratic taste for characteristic k. In particular, parameterizes tastes for minivans and station wagons as:

$$\gamma_{i,mi} = \gamma_{mi} \ln(fs_i)\nu_{i,fv}$$
$$\gamma_{i,sw} = \gamma_{sw} \ln(fs_i)\nu_{i,fv}$$

where fs_i is family size and there is a common preference for fv (family vehicles).

Petrin uses two paricular micromoments: the probability a family purchases any new vehicle given their income, and the average family size of a consumer conditional on purchasing a particular car type (minivan, station wagon, SUV, or full-size van).

The data relies on the CEX automobile supplement: 2660 new vehicle purchases over 6 year period with income and family size for vehicles of interest. Uses supply-side pricing moments as in BLP.

TABLE VII

MEDIAN OWN AND CROSS-PRICE ELASTICITIES[a]

#	Brand	Corn Flakes	Frosted Flakes	Rice Krispies	Froot Loops	Cheerios	Total	Lucky Charms	P Raisin Bran	CapN Crunch	Shredded Wheat
1	K Corn Flakes	−3.379	0.212	0.197	0.014	0.202	0.097	0.012	0.013	0.038	0.028
2	K Raisin Bran	0.036	0.046	0.079	0.043	0.145	0.043	0.037	0.057	0.050	0.040
3	K Frosted Flakes	0.151	−3.137	0.105	0.069	0.129	0.079	0.061	0.013	0.138	0.023
4	K Rice Krispies	0.195	0.144	−3.231	0.031	0.241	0.087	0.026	0.031	0.055	0.046
5	K Frosted Mini Wheats	0.014	0.024	0.052	0.043	0.105	0.028	0.038	0.054	0.045	0.033
6	K Froot Loops	0.019	0.131	0.042	−2.340	0.072	0.025	0.107	0.027	0.149	0.020
7	K Special K	0.114	0.124	0.105	0.021	0.153	0.151	0.019	0.021	0.035	0.035
8	K Crispix	0.077	0.086	0.114	0.034	0.181	0.085	0.030	0.037	0.048	0.043
9	K Corn Pops	0.013	0.109	0.034	0.113	0.058	0.025	0.098	0.024	0.127	0.016
10	GM Cheerios	0.127	0.111	0.152	0.034	−3.663	0.085	0.030	0.037	0.056	0.050
11	GM Honey Nut Cheerios	0.033	0.192	0.058	0.123	0.094	0.034	0.107	0.026	0.162	0.024
12	GM Wheaties	0.242	0.169	0.175	0.025	0.240	0.113	0.021	0.026	0.050	0.043
13	GM Total	0.096	0.108	0.087	0.018	0.131	−2.889	0.017	0.017	0.029	0.029
14	GM Lucky Charms	0.019	0.131	0.041	0.124	0.073	0.026	−2.536	0.027	0.147	0.020
15	GM Trix	0.012	0.103	0.031	0.109	0.056	0.026	0.096	0.024	0.123	0.016
16	GM Raisin Nut	0.013	0.025	0.042	0.035	0.089	0.040	0.031	0.046	0.036	0.027
17	GM Cinnamon Toast Crunch	0.026	0.164	0.049	0.119	0.089	0.035	0.102	0.026	0.151	0.022
18	GM Kix	0.050	0.279	0.070	0.101	0.106	0.056	0.088	0.030	0.149	0.025
19	P Raisin Bran	0.027	0.037	0.068	0.044	0.127	0.035	0.038	−2.496	0.049	0.036
20	P Grape Nuts	0.037	0.049	0.088	0.042	0.165	0.050	0.037	0.051	0.052	0.047
21	P Honey Bunches of Oats	0.100	0.098	0.104	0.022	0.172	0.109	0.020	0.024	0.038	0.033
22	Q 100% Natural	0.013	0.021	0.046	0.042	0.103	0.029	0.036	0.052	0.046	0.029
23	Q Life	0.077	0.328	0.091	0.114	0.137	0.046	0.096	0.023	0.182	0.029
24	Q CapN Crunch	0.043	0.218	0.064	0.124	0.101	0.034	0.106	0.026	−2.277	0.024
25	N Shredded Wheat	0.076	0.082	0.124	0.037	0.210	0.076	0.034	0.044	0.054	−4.252
26	Outside good	0.141	0.078	0.084	0.022	0.104	0.041	0.018	0.021	0.033	0.021

[a] Cell entries i, j, where i indexes row and j column, give the percent change in market share of brand i with a one percent change in price of j. Each entry represents the median of the elasticities from the 1124 markets. The full matrix and 95% confidence intervals for the above numbers are available from http://elsa.berkeley.edu/~ nevo.

Figure 19: Estimated Cross-Price Elasticities.

TABLE 4
Parameter Estimates for the Demand-Side Equation

Variable	OLS Logit (1)	Instrumental Variable Logit (2)	Random Coefficients (3)	Random Coefficients and Microdata (4)
	A. Price Coefficients (α's)			
α_1	.07	.13	4.92	7.52
	(.01)**	(.01)**	(9.78)	(1.24)**
α_2			11.89	31.13
			(21.41)	(4.07)**
α_3			37.92	34.49
			(18.64)**	(2.56)**
	B. Base Coefficients (β's)			
Constant	−10.03	−10.04	−12.74	−15.67
	(.32)**	(.34)**	(5.65)**	(4.39)**
Horsepower/weight	1.48	3.78	3.40	−2.83
	(.34)**	(.44)**	(39.79)	(8.16)
Size	3.17	3.25	4.60	4.80
	(.26)**	(.27)**	(24.64)	(3.57)*
Air conditioning standard	−.20	.21	−1.97	3.88
	(.06)**	(.08)**	(2.23)	(2.21)*
Miles/dollar	.18	.05	−.54	−15.79
	(.06)**	(.07)	(3.40)	(.87)**
Front wheel drive	.32	.15	−5.24	−12.32
	(.05)**	(.06)**	(3.09)	(2.36)**
Minivan	.09	−.10	−4.34	−5.65
	(.14)	(.15)	(13.16)	(.68)**
Station wagon	−1.12	−1.12	−20.52	−1.31
	(.06)**	(.07)**	(36.17)	(.36)**
Sport-utility	−.41	−.61	−3.10	−4.38
	(.09)**	(.10)**	(10.76)	(.41)**
Full-size van	−1.73	−1.89	−28.54	−5.26
	(.16)**	(.17)**	(235.51)	(1.30)**
% change GNP	.03	.03	.08	.24
	(.01)**	(.01)**	(.02)**	(.02)**

Note.—Standard errors are in parentheses. A quadratic time trend is included in all specifications.
* Z-statistic >1.
** Z-statistic >2.

Figure 20: Petrin: Parameter Estimates.

TABLE 9
IMPLIED MARKUPS DERIVED FROM DEMAND-SIDE ESTIMATES AND BERTRAND-NASH PRICING ASSUMPTION, 1981–93 (2,407 Models)

Statistic	OLS Logit (1)	Instrumental Variable Logit (2)	Random Coefficients (3)	(4)	Random Coefficients and Microdata (5)	(6)
Median	$13,834	$7,513	$2,593	36.7%	$1,439	15.0%
Mean	$13,904	$7,551	$4,017	40.7%	$1,753	16.7%
10%	$13,647	$7,413	$1,628	27.8%	$819	11.2%
90%	$14,297	$7,765	$8,357	62.6%	$2,856	24.8%
Standard deviation	$257	$140	$4,089	14.0%	$1,229	6.2%
Estimated marginal costs that are negative	73.7%	22.6%	0%		0%	

NOTE.—Percentage markups are estimated markups divided by observed prices. They are not reported for instrumental variable and OLS logits because the estimated marginal cost is negative for many vehicles. Dollars are 1982–84 CPI adjusted.

Figure 21: Petrin: Implied Markups.

TABLE 13
CHANGE IN U.S. WELFARE FROM THE MINIVAN INNOVATION, 1984–88 ($ Millions)

Year	Compensating Variation	Change in Producer Profits	Welfare Change
1984	367.29	−36.68	330.61
1985	625.04	−25.07	599.97
1986	439.93	27.30	467.23
1987	596.59	29.75	626.34
1988	775.70	110.24	885.94
Total	2,804.55	105.54	2,910.09

NOTE.—Computations were done using 1982–84 CPI-adjusted dollars.

Figure 22: Petrin: Welfare Estimates.

8.3 Gentzkow 2007

Attempts to understand whether or not print and online newspapers are substitutes or complements, and to determine the welfare impact of the introduction of the online edition of the Washington Post.

TABLE 4—LINEAR PROBABILITY MODEL OF *POST* CONSUMPTION

	OLS	IV		
		(1)	(2)	(3)
Dependent variable: Read Post *last 5 days*				
Read post.com last 5 days	0.0464**	−0.4132**	−0.4579**	−0.4381**
	(0.0090)	(0.107)	(0.119)	(0.141)
Other Internet news				0.0244
				(0.0195)
Industry controls			X	
Occupation controls			X	
Overidentification test *p* value		0.302	0.431	0.219
R-squared	0.333	0.214	0.207	0.202
N	14313	14313	14313	14313

Notes: Robust standard errors are in parentheses. The first row gives coefficients on a dummy for reading the post.com in the last five weekdays. IV regressions instrument for post.com consumption with dummy variables for Internet access at work, fast Internet connection, and reported use of the Internet for research/education and work-related tasks. Overidentification test *p* value is the *p* value from a standard Sargan test. Other Internet news is a dummy for online news use other than online newspapers. Industry controls are dummies for 12 industry categories. Occupation controls are dummies for 11 occupation categories. All regressions include controls for *Washington Times* readership, age, sex, education (four categories), white-collar work, computer work, employment status, income, political party, date of survey, location of residence within the DMA (six categories), and dummy variables for the number of missing values. The regressions omit observations where print newspapers are not in the choice set (consumer reports that she generally reads no newspaper sections) and control for presence of online newspapers in the choice set (whether the consumer used the Internet in the last 30 days).

** Significant at 1 percent.

Figure 23: Gentzkow: Linear Probability Model.

Issue: identification between complementarity vs. correlated preferences.

Uses variation in accessibility of Internet at work: will affect utility/price of online Post.com, but not of paper version. If goods are complementary, increasing accessibility of Internet should increase demand of paper version of Post; this exclusion restriction means that this is the only channel through which paper demand for Post can go up.

Also panel variation: complementarity makes it unlikely that only one good will be consumed on a given day for a consumer, but with correlated random effects (conditional on a consumer's propensity to consume goods), variation in consumption should be uncorrelated across goods.

Data: Survey (16K adults) in DC DMA between March 2000 - Feb 2003. individual/household chars and enumeration of all local print newspapers read over the past 24 hours and 5 days, and readership of local online newspapers. Wash Post (print + online) and Wash Times. Outside option will be other news sources and non-consumption.

Utility from good j

$$\bar{u}_{ijt} = -\alpha p_j + \delta_j + \boldsymbol{x}_i \boldsymbol{\beta}_j + \nu_{ij} + \tau_{it}$$

and utility from bundle r:

$$u_{irt} = \sum_{j \in r} \bar{u}_{ijt} + \Gamma_r + \epsilon_{irt}$$

Identification of Γ relies on: using whether consumer users Internet at work, uses Internet for work or education related tasks, has high speed internet at home (all included for Post.com utility but excluded from Post print and Times print).

Uses supply side pricing moments to help assist with identification of α.

Estimated with both MSM and SML.

Computes industry outcomes and changes in producer and consumer surplus if Post.com is removed.

TABLE 5—PARAMETER ESTIMATES FROM FULL MODEL: OBSERVABLE CHARACTERISTICS

	Post	post.com	*Times*
Age	0.661**	−0.545**	0.688**
	(0.0407)	(0.0628)	(0.122)
Female	−0.473**	−0.423**	−3.20**
	(0.0921)	(0.148)	(0.344)
High school	1.95**	2.87**	1.29
	(0.214)	(0.633)	(0.856)
College	2.54**	4.03**	1.49
	(0.237)	(0.654)	(0.928)
Grad school	2.76**	4.16**	1.19
	(0.252)	(0.670)	(0.964)
Computer job	−0.567*	1.22**	0.322
	(0.223)	(0.323)	(0.682)
White-collar job	−0.447**	0.431*	−0.591
	(0.118)	(0.195)	(0.423)
Full-time	−1.13**	0.935**	−0.044
	(0.141)	(0.227)	(0.450)
Log income	0.709**	0.217	0.934**
	(0.0729)	(0.118)	(0.260)
Democrat	0.217**	0.326*	−0.027
	(0.099)	(0.163)	(0.346)
Republican	−0.193	−0.0538	2.902**
	(0.119)	(0.193)	(0.408)
Constant	6.97**	−1.54	−8.85**
	(0.852)	(1.18)	(1.91)
N	16179	16179	16179

Notes: Standard errors in parentheses. Details of the model are given in the text. Age is measured in units of ten years. High school, college, and graduate school are mutually exclusive categories. Computer and white-collar job are dummies for reported occupations in these categories and are also mutually exclusive. Full-time is a dummy for full-time employment. Democrat and Republican indicate registered members of the parties. Additional model parameters are shown in Table 6. Not shown in any table are dummies for the number of missing observations, location of the respondent's residence within DC, time dummies (in six-month intervals), and having print and online papers in the choice set.

* Significant at 5 percent.

Figure 24: Gentzkow: Parameter Estimates.

TABLE 6—PARAMETER ESTIMATES FROM FULL MODEL: OTHER

Interaction terms		Excluded variables (coefficient in utility of post.com)	
Post-post.com	−1.285**	Internet at work	1.357**
	(0.2307)		(0.180)
Post-Times	0.0809	Fast connection	0.146
	(0.2479)		(0.193)
post.com-*Times*	−1.231**	Use for education-related	0.361
	(0.4832)		(0.212)
Nonlinear parameters		Use for work	0.582**
τ	6.846**		(0.222)
	(0.5027)		
γ	0.0454**		
	(0.0179)		

Notes: Standard errors in parentheses. The model also includes a third-order interaction term for the *Post-Times*-post.com bundle which is not reported in the table. Fast connection indicates consumers with DSL, cable modem, or T1 connections at home. Use variables were responses to the question, "In what ways do you use online services?"
** Significant at 1 percent.

Figure 25: Gentzkow: Parameter Estimates.

TABLE 8—IMPACT OF THE ONLINE EDITION ON DEMAND FOR PRINT

Case 1: Full model	
Cross-price derivative	8,358
	(1,436)
Change in print readership	−26,822
	(4,483)
Change in print profits	−$ 5,466,846
	(913,699)
Case 2: Model with observable characteristics only	
Cross-price derivative	−8,421
	(752)
Change in print readership	25,655
	(2,270)
Change in print profits	$ 5,229,009
	(462,771)
Case 3: Model with no heterogeneity	
Cross-price derivative	−16,143
	(702)
Change in print readership	51,897
	(2,254)
Change in print profits	$10,577,720
	(459,464)

Notes: Standard errors in parentheses. The table shows three measures of the online edition's impact. The cross-price derivative is the change in post.com readership when the *Post*'s price is increased by $.10. Change in print readership and print profits are the total changes for the *Post* when the online edition is added to the choice set. The table shows the estimated values in three models. Case 1 is the estimates from the full model. Case 2 is a model with observable consumer characteristics but no unobservables other than the i.i.d. logit errors. Case 3 is a model with no observable or unobservable consumer heterogeneity except the i.i.d. logit errors.

Figure 26: Gentzkow: Counterfactual Estimates.