Demand Systems for Empirical Work in IO*

John Asker, Allan Collard-Wexler

October 5, 2007

Overview

Demand systems often form the bedrock upon which empirical work in industrial organization rest. The next 2.5 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 start by talking about the different instances in which demand estimation arises and the core problems we face when estimating demand. After reviewing standard data forms, I will then 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 endogeniety 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.

*These notes draw from a variety of sources, in particular Ariel Pakes' lecture notes from when I was a grad student. I have rewritten large amounts so any mistakes are mine.

Why spend time on Demand Systems?

- In IO we usually care about care about comparative statics of one form or another. Usually demand is important to this: Think about pre and post merger pricing, tax incidence, monopoly vs duopoly pricing, effect of predative pricing policies etc.

- Also care about welfare impacts: need a well specified demand system for welfare calculations

- 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 adverts or advertising as a barrier to entry.

- Understanding the cross-price elasticities of good is often crucial to "preliminary" issues in policy work, such as market definition in antitrust cases. We will talk about the antitrust applications of demand models in the third lecture. Note that this is the largest consumer of Ph.D's in Empirical I.O. by a long shot!

- Determinants of Market Power: should we allow two firms to merge? Is there collusion going on in this industry (unusually large markups)?

- Determinants of Innovation: once you have markups you know which products a firm will want to produce (SUV's, cancer drugs instead of malaria treatments).

- Value of Innovation: compute consumer surplus from the introduction of a new good (minivans and CAT scan).

- The tools used in demand estimation are starting to be applied in a variety of other contexts to confront empirical issues, of there is likely to be some intellectual arbitrage for your future research.

Data...

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...).
- 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.
- Note that you have a lot of information here: 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).
- So can fill in a lot of blanks
- 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.

Example: Autos

Bresnahan 1987: Competition and Collusion in 1950s Auto Market

Wanted to examine the hypothesis that the dramatic decrease in the price of Autos in 1955 was due to the temporary breakdown of a collusive agreement. 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 Auto. He has about 85 models over 3 years.

The "magic" in these approaches is using demand data to back out marginal costs, *without any cost data.*

Question: What are the empirical modelling issues here?

Approaches to demand estimation

Approaches breakdown along the following lines:

- single vs multi-products
- representative agent vs heterogenous agent
- within multi-product: whether you use a product space or characteristic space approach

Revision: Single Product Demand Estimation

- Start with one homogenous product.
- Assume an isolelastic demand curve for product j in market t:

$$\ln(q_{jt}) = \alpha_j p_{jt} + X_{jt}\beta + \xi_{jt} \tag{1}$$

Note that price elasticity $\eta_{jjt} = \alpha_j p_{jt}$.

- X_{jt} could just be an intercept for now (constant term).
- ξ_{jt} are unobserved components of demand (often called unobserved product quality).

Let's go to the supply side for a second since the firm selling product j is presumably allowed to choose it's price (if we are in an unregulated market).

Firms get to choose prices. The pricing rule of a monopolist is to maximize profits:

$$\pi_{jt} = (p_{jt} - c_{jt})q_{jt} \tag{2}$$

(assuming constant marginal costs for now)

The F.O.C. for this problem is:

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

Problem 1: Endogeneity of Prices

- 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:

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

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

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

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 (if I don't have the data, why should a firm have this data)!
- Remember that these guys spend their lives thinking about pricing.
- Moreover, won't firms react if they see higher than expected demand yesterday?
- <u>Note:</u> 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".

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

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

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 $E[XZ] \neq 0$.

This is 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 $\alpha_i \sim F\left(\alpha|\theta\right)$ and $\gamma_i \sim G\left(\alpha|\tau\right)$ 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)

Multi-product Systems

Now let's think of a multiproduct demand system to capture the fact that most products have substitutes for each other.

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

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.

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.
  - Can get around this to some extent by imposing more structure in the form of 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.

- hard 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.

Product Space Approaches: AIDS Models

I will spend more than an average amount of time on AIDS (Almost Ideal Demand System, which was published in 1980 by Deaton and Mueller AER and 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.

	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. 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.

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).

AIDS 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: expenditure functions look like

$$\log\left(e(w,p)\right) = a\left(p\right) + wb\left(p\right)$$

where w is just a weight between zero and one, a is a quadratic function in p, and b is a power function in p. Using Shepards Lemma we can get shares of expenditure within groups as:

$$w_i = \alpha_i + \sum_j \theta_{ij} \log\left(p_j\right) + \beta_i \log\left(\frac{x}{P}\right)$$

where x is expenditure on the group, P is a price index for the group and everything else should be self explanatory. Dealing with the price index can be a pain. There are two ways that are used. One is the "proper" specification

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

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

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

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.

Example: Hausman 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:

$$w_i = \alpha_i + \sum_j \theta_{ij} \log(p_j) + \beta_i \log\left(\frac{x}{P}\right) + \varepsilon$$

[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

$$\log q_m = \beta_m \log y_B + \sum_k \sigma_k \log(\pi_k) + \alpha_m + \varepsilon$$

where q_m is the segment quantity purchased, y_B 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.

$$\log u_t(\text{Beer Spending}) = \beta_0 + \beta_1 \log y_t(\text{Income}) + \beta_2 \log P_B(\text{Price Index for Beer}) + Z_t\delta + \varepsilon$$

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:

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

  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). 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 1: 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)

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

Note: Symmetry imposed during estimation.

Figure 2: 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 (0.072)	0.464 (0.060)	0.397 (0.039)	0.254 (0.043)	0.201 (0.037)
Coors Light	0.569 (0.085)	−4.598 (0.115)	0.407 (0.058)	0.452 (0.075)	0.482 (0.061)
Old Milwaukee Light	1.233 (0.121)	0.956 (0.132)	−6.097 (0.140)	0.841 (0.112)	0.565 (0.087)
Lite	0.509 (0.095)	0.737 (0.122)	0.587 (0.079)	−5.039 (0.141)	0.577 (0.083)
Molson Light	0.683 (0.124)	1.213 (0.149)	0.611 (0.093)	0.893 (0.125)	−5.841 (0.148)

Figure 3: Segment 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

Figure 4: Overall Elasticities

Chaudhuri, Goldberg and Jia Paper

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 usefull 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? [1]

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

[1] 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.

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.

- 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 (not if they are licensed or not!).

- Some products enter and exit the sample. How can the AIDS model deal with this?

- Some products have different dosages than others. How does one construct quantity for this market.

Results

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

- 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. $p = \frac{mc}{1+1/\eta_j j}$. 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).

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 5: Summary Statistics

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 6: Marginal Costs

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

	Elasticity with respect to:							
	Prices of foreign product groups			Prices of domestic product groups				Overall quinolones expenditure
Product group	Cipro	Norflo	Oflo	Cipro	Norflo	Oflo	Sparflo	
Foreign ciprofloxacin	−5.57*	$-0.13^{\dagger}$	−0.15*	4.01*	$0.11^{\dagger}$	$0.11^{\dagger}$	0.16*	1.37*
	(1.79)	(0.07)	(0.07)	(1.84)	(0.06)	(0.06)	(0.06)	(0.29)
Foreign norfloxacin	$-4.27^{\dagger}$	−0.45	$-4.27^{\dagger}$	$3.50^{\dagger}$	−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*	$-0.10^{\dagger}$	−1.38*	−0.09	$0.09^{\dagger}$	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 7: Elasticity Estimates

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.

Formal Treatment

- Utility of the individual:

  $$U_{ij} = U\left(x_j, p_j, v_i; \theta\right)$$

  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

  $$U_{ij} > U_{ik}\, \forall k \quad \text{[note that all preference relations are assumed to be strict]} \tag{4}$$

- 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 $s_j(\mathbf{x}, \mathbf{p}|\theta)$.

- 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.

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.

Examples

Anderson, de Palma and Thisse go through many of these in very close detail...

Horozontally Differentiated vs **Vertically Differentiated** - Horozontally differentiated means that, setting aside price, people dissagree 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 icecream 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 II $[F(\epsilon) = e - e^{-\epsilon}]$. This is a very helpful assumption as it allows for the aggregate shares to have an analytical form.
  - 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.
  - 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 $j \in S(j)$.
- Probabilities are given by:

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

For two different goods in different segments, the relative choice probabilities are:

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

- 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.

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

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!

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.

1. In the vertical model people agree on the relative quality of products, hence there is a clear ranking of products in terms of quality

2. The only difference between people is that some have less willingness to pay for quality than others

3. Hence (recall) utility will look like

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

4. To gain the shares predicted by the model we need to

5. 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.)

6. The lowest good is the outside good (good 0) - we normalise this to zero ($\overline{u} = 0$)

7. Choose 0 if

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

8. 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

9. 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 $j = 1, ...J$.

10. Question: What parameters are identified in θ? What are the sources of identification for each parameter?

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

$$\max_\theta \quad n \sum_j s_j^0 \log [s_j(\theta)]$$

Where n_j is the count of individuals choosing the object.

Another Example: Logit

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

Here the utility is

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

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

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.

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 derivative is $\frac{\partial s}{\partial p} = -s(1 - s)$. That is, the own price derivative only depends on shares, which in turn means that if we see two products with the same share, they must have the same mark-up, under most pricing models.

2. Cross-price elasticities are $s_j s_k$. This means that the substitution matrix is solely a function of shares and not relative proximity of products in characteristic space. This is a bit crazy for products like 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).

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

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

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

$$\ln[s_j] - \ln[s_0] = \delta_j = \sum_k \beta_k x_{kj} - \alpha p + \xi_j$$

$$\xi_j = \ln[s_j] - \ln[s_0] - \left(\sum_k \beta_k x_{kj} - \alpha p + \xi_j\right)$$

-
  - 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.

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:

Figure 8: The Reliant Regal

- 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.

- 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.

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
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.

[Quickly review OLS as reminder]

Recall that

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

so, once we know δ_j we can run this regression as OLS or IV or whatever and get the ξ_j. Then we can construct the moments for the estimator.

So how to do this? This is the bit where BLP does something very cool.

- 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). Neither, Nevo or BLP point this out, although both exploit it, but the following is also a contraction

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

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

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

- 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.

In evaluating the moment restrictions we impose a norm which will usually be the L^2 norm. ie

$$\|G_{J,ns}\|_2 = \sqrt{\sum_j \left[\xi_j f_j(x, w)\right]^2}$$

Step 4: 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. Nevo uses the difference between linear and non-linear parameters to estimate brand fixed-effects (and I've done brand/market fixed effects too).

- 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, Guass etc to code it up. I like Matlab, personally.
  - There are different ways to numerically search for minimums: always prefer use a simplex search algorithm over derivative based methods, simplex searches take a little longer but are more robust to poorly behaved functions. Also start you code from several different places before believing a given set of results. [in matlab fminsearch is a good tool].
  - 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. My favorite minimizers 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...

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).

Nevo

- 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.

- 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:

$$\Pi_F(j) = \sum_{j \in F(j)} s_{jt}(p_{jt} - mc_{jt}) \tag{5}$$

Taking the first-order condition you get:

$$\frac{\partial \Pi}{\partial p_{jt}} = s_{jt} - \sum_{k \in F(j)} \frac{s_{kt}}{p_{kt}}(p_{jt} - mc_{jt}) = 0 \tag{6}$$

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:

$$\mathbf{s} + \mathbf{\Omega} \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}}(\mathbf{p} - \mathbf{c}) = 0 \tag{7}$$

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

$$\mathbf{c} = p + (\mathbf{\Omega} \frac{\partial \mathbf{s}}{\partial \mathbf{p}})^{-1}\mathbf{s} \tag{8}$$

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

$$\ln(mc_{jt}) = X_{jt}\delta + \omega_{jt} \tag{9}$$

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.

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 9: Costs and Back of the envellope markups in the RTE Cereal Industry.

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 10: Estimated BLP Model.

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 11: Estimated Cross-Price Elasticities.

Overview of BLP

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

year/model is an observation = 2217 obs

Petrin

I will just spend a short time talking about Amil Petrin's paper.

- We often want to quantify the benifits of innovation.
- You need a demand curve to do this since we want to know what people would be willing to pay above the market price.
- 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.

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 12: BLP Model Estimates.

One important innovation is the use of "Micro-Moments". Moment conditions coming from micro-data. So for example, one might have data coming from the CEX on the average amount of money spend 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:

$$s_t|I < 10000, \theta = \sum_{j>0} \frac{1}{1(I_k < 10000)} \sum_k (s_{ijt}(\theta) 1(I_k < 10000)) \tag{10}$$

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.

The second important point in Petrin is how to look at the benifit of a new product. Consumer Surplus in the logit model is given by:

$$u_{\tilde{j}t} = X_{\tilde{j}t}\beta - \alpha p_{\tilde{j}t} + \xi_{\tilde{j}t} + \epsilon_{i\tilde{j}t}(11)$$

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 13: Elasticities from the BLP Model.

Given estimates of the value of characteristics, and some assumption on where the ξ's are coming from we can compute the utility of the new good and the new set of prices in the market. We also need an estimate of marginal cost of the new product which could come from our supply side moment. But how do we compare consumer surplus before and after the introduction of the new product, since there is now some extra $\epsilon_{i\tilde{j}t}$ on the table that did not exist before.

The formula for consumer surplus in the logit model is given by (which depends on the value of ϵ_{ijt} given that I choose a specific product, i.e. $E(\epsilon_{ijt}|u_{ijt} > u_{ikt} \forall k)$):

$$CS = \frac{1}{\alpha} \sum_j \ln(\sum_j \exp(\delta_{jt})) + C \tag{12}$$

where δ_{jt} is good old mean utility term. So the change in consumer surplus would be:

$$\Delta CS = \frac{1}{\alpha} [\sum_{\tilde{j}} \ln(\sum_{\tilde{j}} \exp(\delta_{\tilde{j}t})) - \sum_j \ln(\sum_j \exp(\delta_{jt}))] \tag{13}$$

Note that there are two effects here. The first pertains to the change in δ's when the new product is introduced. The second has to do with a change in the number of products over which you take this sum. Note that if I double the number of products by relabeling them I can get a big increase in consumer welfare!

Question Where is this effect coming from that relabeling products makes you better off. It has to do with $E(\epsilon_{ijt}|u_{ijt} > u_{ikt} \forall k))$ and correlation...

TABLE 5
RANDOM COEFFICIENT PARAMETER ESTIMATES

	Random Coefficients (γ's)	
Variable	Uses No Microdata (1)	Uses CEX Microdata (2)
Constant	1.46	3.23
	(.87)*	(.72)**
Horsepower/weight	.10	4.43
	(14.15)	(1.60)**
Size	.14	.46
	(8.60)	(1.07)
Air conditioning standard	.95	.01
	(.55)*	(.78)
Miles/dollar	.04	2.58
	(1.22)	(.14)**
Front wheel drive	1.61	4.42
	(.78)**	(.79)**
γ_{mi}	.97	.57
	(2.62)	(.10)**
γ_{sw}	3.43	.28
	(5.39)	(.09)**
γ_{su}	.59	.31
	(2.84)	(.09)**
γ_{pv}	4.24	.42
	(32.23)	(.21)**

NOTE.—The OLS and instrumental variable models assume that these random coefficients are zero. Standard errors are in parentheses. A quadratic time trend is included in all specifications. The subscript *mi* stands for minivan, *sw* for station wagon, *su* for sport-utility, and *pv* for full-size passenger van.
* Z-statistic >1.
** Z-statistic >2.

Figure 14: Random-Coefficients for Petrin paper, with and without micro-moments.

Merger Analysis

I will spend a short time on the mechanics of merger predictions, which are discussed in Nevo's RAND paper and Hausman, Leonard and Zona's AES paper.

Remember from the supply side that the FOC implies:

$$\mathbf{c} = \mathbf{p} + (\boldsymbol{\Omega} \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}})^{-1}\mathbf{s} \tag{14}$$

What happens when two firms decide to merge? They start caring about the effect of the price of one firm on the market share of the other firm's products. Formally, the ownership matrix Ω now has to account for the new pattern in the industry, which we now call Ω^* .

We need to find a new set of prices $\mathbf{p_t} = (_{1t}, ..., _{Jt})$ which satisfies the FOC of the firms:

$$\mathbf{p_t} = \mathbf{c_t} - (\boldsymbol{\Omega}^* \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}}(\mathbf{p_t}))^{-1}\mathbf{s}(\mathbf{p_t}) \tag{15}$$

So we need to find the root of a system of equations with J elements. Note that we can either get the expressions for $\mathbf{s}(\mathbf{p_t})$ and $\frac{\partial \mathbf{s}}{\partial \mathbf{p}}(\mathbf{p_t})$ from the logit, BLP or AIDS models to do merger analysis.

An important question here is how well are these merger models fitting the data. One way to examine this problem is to look at mergers that occured and see how closely the realized price changes match the predictions of the Model. Peters (2006) and Whinston (2006) have some pretty depressing results on this stuff.

TABLE 7

Equilibrium Prices with and without the Minivan, 1984: 1982–84 CPI-Adjusted Dollars

	Price			%
	With Minivan	Without Minivan	ΔPrice	ΔPrice
	A. Largest Price Decreases on Entry			
GM Oldsmobile Toronado (large sedan)	15,502	15,643	−141	.90
GM Buick Riviera (large sedan)	15,379	15,519	−139	.89
GM Buick Electra (large sedan)	12,843	12,978	−135	1.04
GM Chevrolet Celebrity (station wagon)	8,304	8,431	−127	1.51
Ford Cadillac Eldorado (large sedan)	19,578	19,704	−126	.64
Ford Cadillac Seville (large sedan)	21,625	21,749	−125	.57
GM Pontiac 6000 (station wagon)	9,273	9,397	−123	1.31
GM Oldsmobile Ciera (station wagon)	9,591	9,714	−123	1.27
GM Buick Century (station wagon)	8,935	9,056	−121	1.34
GM Oldsmobile Firenza (station wagon)	7,595	7,699	−104	1.35
	B. Largest Price Increases on Entry			
Chrysler LeBaron (station wagon)	9,869	9,572	297	3.10
Volkswagen Quattro (station wagon)	13,263	13,079	184	1.41
Chrysler (Dodge) Aries K (station wagon)	7,829	7,659	170	2.22
AMC Eagle (station wagon)	10,178	10,069	109	1.08

Note.—Equilibrium prices without minivans are estimated using the model with microdata and Bertrand-Nash first-order conditions. Bertrand-Nash pricing with random coefficients does not a priori determine signs of firm-specific price changes.

Figure 15: Prices before and after introduction of Minivan.

TABLE 5 Predicted Percent Change in Prices and Quantities as a Result of Mergers

	Post and Nabisco		GM and Nabisco		GM and Chex		Kellogg and Quaker Oats		GM and Quaker Oats	
	p	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*
K Corn Flakes	.0	.0	.0	.1	.0	.1	.0	.5	.0	.3
K Raisin Bran	.1	.1	.1	.3	.1	.2	1.4	−1.7	.5	.7
K Frosted Flakes	.0	.0	.0	.1	.0	.1	.3	−.4	.1	.3
K Rice Krispies	.0	.1	.1	.2	.1	.4	5.1	−4.1	.7	2.0
K Frosted Mini Wheats	.0	.2	.0	.2	.1	.3	2.7	−4.1	.3	2.9
K Froot Loops	.0	.1	0	.2	.1	.5	9.3	−15.3	.7	8.0
K Special K	.0	.1	.1	.2	.0	.2	.2	.2	.1	.4
K NutriGrain	.0	.0	.1	.1	.0	.1	.0	.4	.1	.3
K Crispix	.0	.1	.0	.2	.1	.4	3.4	−3.8	.5	2.7
K Cracklin Oat Bran	.0	.1	.0	.2	.0	.4	3.4	−6.8	.4	3.7
GM Cheerios	.0	.2	.7	−.9	1.1	−1.3	.5	1.3	4.1	−3.5
GM Honey Nut Cheerios	.0	.1	.5	−.6	.8	−.9	1.0	3.2	11.5	−11.2
GM Wheaties	.0	.0	.0	0	.1	−.1	.1	.5	.1	.3
GM Total	.0	.1	.3	−.8	.2	−.6	.1	.4	.2	.1
GM Lucky Charms	.0	.1	.3	−.4	.7	−.8	.8	3.3	9.3	−10.6
GM Trix	.0	.1	.3	−.3	.7	−.9	.7	3.5	8.6	−9.6
GM Raisin Nut	.0	.2	.4	−.7	.5	−.9	.3	1.5	1.8	−2.7
P Raisin Bran	.9	−1.5	.0	.5	.0	.4	.1	1.5	.2	1.7
P Grape Nuts	1.5	−2.8	.1	.7	.0	.4	.1	2.3	.1	3.0
Q 100% Natural	.0	.1	.0	.3	.0	.5	10.2	−17.0	11.4	−19.3
Q Life	.0	.1	.0	.3	.1	.5	15.5	−16.7	23.8	−25.3
Q CapNCrunch	.0	.1	.0	.3	.1	.4	16.8	−16.7	29.1	−30.9
R Chex	.0	.2	.0	.3	12.2	−19.0	.0	2.1	.1	3.4
N Shredded Wheat	3.1	−8.6	7.5	−18.8	.0	.4	.0	1.9	.0	2.5

Figures are the median change for each brand over the 45 cities in the last quarter of 1992, and are based on Table 2.

Figure 16: Predicted Price Changes in the Nevo Model.