

Demand Systems in Industrial Organization: Part II*

John Asker

February 16, 2015

1 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.

1.1 Formal Treatment

- Utility of the individual:

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

for $j = \{0, 1, 2, 3, \dots, J\}$.

- 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

*These notes draw from a variety of sources: in particular Ariel Pakes' lecture notes from grad school days, and from (co-teaching with) Robin Lee and Allan Collard-Wexler.

- 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

$$U_{ij} > U_{ik} \quad \forall k \quad [\text{note that all preference relations are assumed to be strict}] \quad (1)$$

- This means that the set of consumers that choose good j is given by

$$\mathbb{S}_j(\theta) = \{v | U_{ij} > U_{ik} \quad \forall k\}$$

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.

1.1.1 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 to interpret our coefficients and do estimation.

1.2 Examples (Briefly)

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

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

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

where the $(\delta_j - \nu_i)^2$ 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

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

- 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

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

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

I.e.:

$$Pr(U_{ij} \geq U_{ik} \forall k) = \frac{\exp(\delta_j)}{\sum_{k=0, \dots, J} \exp(\delta_k)} \quad (2)$$

- 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:

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

$\lambda_k \in (0, 1]$ 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:

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

- 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 need 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.

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

1.3 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!

1.3.1 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

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

- 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 ($\bar{u} = 0$)
 3. Choose 0 if

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

4. Hence $S_0 = \left\{ \nu \mid \nu > \frac{\delta_1}{p_1} \right\}$. Thus if ν is distributed lognormally, $\nu = \exp(\sigma x + \mu)$ where x is distributed standard normal, then choose 0 if

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

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 $j = 1, \dots, J$.

- 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

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

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 $\epsilon_{ij} \stackrel{iid}{\sim}$ 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:

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

$$s_0 = \frac{1}{1 + \sum_{q \geq 1} \exp[\delta_q - p_q]}$$

1.4 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 $\theta = \{\delta_1, \dots, \delta_J, \mu, \sigma\}$). (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 than the AIDS model since there are only $J + 2$ parameters to estimate versus $J^2 + J$ of them.

1.5 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 endogeneity via simultaneity. This leads to the same economically stupid results that we see in single product demand estimation that ignores endogeneity (like upward sloping demand etc).

1.6 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 variation due 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 $s_j^o = s_j(\xi_1, \dots, \xi_J | \theta)$

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

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

- 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 endogeneity 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.

2 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 endogeneity 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

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

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

ϵ_{ij} : 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

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

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)

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

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.
- 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). **If you have micro data it is almost always a very good idea to use it.**
- 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.



Figure 1: The Reliant Regal

2.1 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.

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

Step 0: Search over θ Generally, we are looking for $\theta \equiv \{\lambda, \beta\}$ 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 θ .

For a given evaluation of the objective function and a given θ , 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 $\xi(\theta)$ that, along with our candidate parameter vector (θ), induces our model to match the shares in the data, and then compute the GMM objective $G(\xi(\theta), \mathbf{Z}; \theta)$ where \mathbf{Z} represent instruments.

Thus, one performs the following steps:

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

- After integrating out the ϵ_{ij} (recall that these are familiar logit errors) the equation for the share is

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

- 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

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

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.
- There 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

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

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

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

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 \rightarrow 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)

2.2 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).

2.3 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:

$$\Pi_F(j) = M \sum_{j \in F(j)} s_{jt}(p_{jt} - mc_{jt}) \quad (3)$$

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

$$\frac{\partial \Pi}{\partial p_{jt}} = s_{jt} + \sum_{k \in F(j)} \frac{s_{kt}}{p_{jt}} (p_{kt} - mc_{kt}) = 0 \quad \forall j \quad (4)$$

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 across all products j in market t to get:

$$\mathbf{s} + \Omega \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}} (\mathbf{p} - \mathbf{mc}) = 0 \quad (5)$$

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

$$\mathbf{mc} = \mathbf{p} + (\Omega \cdot * \frac{\partial \mathbf{s}}{\partial \mathbf{p}})^{-1} \mathbf{s} \quad (6)$$

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

$$\ln(mc_{jt}) = X_{jt}\gamma + \omega_{jt} \quad (7)$$

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 (6) is a function of θ , we can recover an $\omega(\theta)$ during our estimation routine (once we add γ to the parameters being estimated), and we can add $E(\omega(\theta)Z) = 0$ to the previous moment conditions.

2.4 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

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.

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)	<i>Constant</i>	-7.061	0.941	-7.304	0.746
	<i>HP/Weight</i>	2.883	2.019	2.185	0.896
	<i>Air</i>	1.521	0.891	0.579	0.632
	<i>MP\$</i>	-0.122	0.320	-0.049	0.164
	<i>Size</i>	3.460	0.610	2.604	0.285
Std. Deviations (σ_{β} 's)	<i>Constant</i>	3.612	1.485	2.009	1.017
	<i>HP/Weight</i>	4.628	1.885	1.586	1.186
	<i>Air</i>	1.818	1.695	1.215	1.149
	<i>MP\$</i>	1.050	0.272	0.670	0.168
	<i>Size</i>	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					
	<i>Constant</i>	0.952	0.194	0.726	0.285
	$\ln(HP/Weight)$	0.477	0.056	0.313	0.071
	<i>Air</i>	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
	<i>Trend</i>	0.019	0.002	0.026	0.004
	$\ln(q)$			-0.387	0.029

Figure 2: BLP Model Estimates.

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
735i	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 3: 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 735i	90.883	10.101

Figure 4: BLP: Comparison b/w RC and Logit.

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 735i	\$37,490	\$10,975	\$114,802

Figure 5: BLP: Markups.