Demand for Differentiated Products

Advanced Microeconometrics

Instructor: Cristián Sánchez

Spring 2022

Models of Demand for Differentiated Products

- Most of the demand literature focuses on models where consumers choose between substitute products with different characteristics.
- There are two types of differentiation:
  - VERTICAL: everyone agrees what makes a good better; may disagree about how much they value it.
  - HORIZONTAL: people disagree about whether more or less of an attribute is more valuable.
- a classic vertical model would specify utility as

$$u_i(j) = \theta_i q_j - p_j, \quad \theta > 0 \quad \text{for all } i$$

- a classic horizontal model might specify utility as

$$u_i(j) = v - p_j - t|d_{ij}|$$

  where people and products are arranged on a line or on a circle.
- What are good examples of each type of characteristic?

Differentiated Product Demand

- The key issue is that the quantity demanded of one product may/will depend on the **prices and characteristics of all products**.
  - These won't predict things perfectly at either the individual or the aggregate level so we also need to introduce some errors.
  - In a structural model we will think about these errors as **characteristics** or **preferences**.
- Ideally we want to represent demand in a way that is:
  - consistent with basic choice theory
  - parsimonious
  - flexible enough to reflect the data
  - identified given available data
  - easy to estimate

Differentiated Product Demand

Two Types of Models

There are two types of models:

1. Continuous choice models, usually based on a representative consumer who consumes some of everything.
2. Discrete choice models, where each consumer consumes at most one unit of one product, but different consumers may make different choices.

Differentiated Product Demand

Our Focus

- We will focus on **discrete choice models**. These models assume that utility is a function of product characteristics, and make up the bulk of the literature, and provide a natural way for thinking about welfare questions.
- But you should be aware they are not the only option, especially when there is a fixed and small number of products.

Discrete Choice Models

Logit, Nested Logit and Random Coefficient Logit

- Logit-based models constitute the bulk of the literature.

- Vital to understand what different models assume.

- The underlying model: assume a consumer has a choice set (including option of buying nothing) and **chooses one unit of the one option** that maximizes her utility

$$U_{ijt} = \alpha_i (y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}$$

  where i is consumer, j is product, y_i is income.

- In demand context, we almost always choose β_i to be the same across the products.
  - Let's say that there are C observed characteristics.

- ε_{ijt} is an idiosyncratic Type I extreme value (logit) error.

- Specification and estimation often varies across **aggregate** (market share) and **individual level** data.

- When we assume the consumer chooses a single product and she gets the utility of that choice we are ruling out some situations:
  - Products are complements, rather than substitutes.
  - I buy multiple products, and maybe optimize their use (e.g., cars).
- There are papers that relax these assumptions while still maintaining a basically similar utility specification.
- Typically however people do not worry about the multiple purchase problem.
  - Instead can view it as a model of daily consumption choices.
  - Of course, there can be differences between **purchases** and **consumption** and we'll return to this issue.

Discrete Choice Models: Differences Between Models

- The key specification differences are whether:
  - There are unobserved product characteristics ξ_{jt} (valued in same way by all).
    - ξ_{jt}s will be structural errors when we have aggregate demand.
    - Focus of endogeneity concerns.
  - Whether α_i differs across consumers.
  - Whether β_i differs across consumers.
    - If so can get horizontal differentiation between consumers based on observable characteristics.
    - Otherwise the only source of horizontal differentiation is through ε_{ijt}.
  - Whether there are correlations in the εs across similar products.

- Richer models are harder to estimate and may stretch identification given real data.

The Multinomial Logit Model

- Often called the conditional logit model in management and business communities.
- Assume $\alpha_i = \alpha$, $\beta_i = \beta$

$$U_{ijt} = \alpha(y_i - p_{jt}) + x_{jt}\beta + \xi_{jt} + \varepsilon_{ijt}$$

- ε is Type I extreme value, IID across options and time.
- Utility of the outside good ($j = 0$, not buying) is αy_i.
- Consumer chooses one unit of the good that gives the highest utility.
- People only make different choices because of their εs.
- Income drops out (it's common to all products)

$$\begin{aligned} U_{ijt} &= x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \varepsilon_{ijt} \\ &= \delta_{jt} + \varepsilon_{ijt} \end{aligned}$$

Multinomial Logit Model

- (quantity) market share for product is

$$s_{kt} \equiv \Pr(\text{choose } k) = \int \ldots \int_{A_k} f(\varepsilon_1, \ldots, \varepsilon_J \mid \delta)\, d\varepsilon_1 \ldots d\varepsilon_J$$

where A_k defines $\varepsilon-$ space where k gives highest utility

$$= \frac{\exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})}{1 + \sum_{j=1}^{J} \exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})} = \frac{e^{\delta_k}}{1 + \sum_{j=1}^{J} e^{\delta_j}}$$

where δ_j is mean utility.

- This is a super-convenient functional form for a high dimensional integral!

Multinomial Logit Model

Estimation with Aggregate Data

- Aggregate data means that we observe the characteristics (incl. price) and quantities sold of each product.
  - We do not observe who buys what.
- However, given that we are assuming everyone is identical, up to the logit error, then this is not a problem.

- In fact, if we assume that the market is large, then we can claim to observe the choice probabilities as market shares (i.e., ignoring sampling error)

$$\begin{aligned} s_{jt} &= \frac{\text{Number of people buying } j \text{ in market } t}{M_t} \\ &\approx \frac{\text{Number of units sold of } j \text{ in market } t}{M_t} \end{aligned}$$

  - With a small sample of individual data, we would only see some peoples' choices.
  - This is less good in the sense that the model predicts choice probabilities.

Multinomial Logit Model

Estimation with Aggregate Data: Market Definition

- But to calculate market shares there is a problem, how do you define a market (i.e., the set of people who might buy)?
  - We avoid this problem with individual data as we see a given set of individuals and we know which of them do not buy anything.[1]
  - With aggregate data, we need a measure of potential purchases to define market shares (including for the outside good).
- In general people just use something rough-and-ready:
  - cereals (one serving per day per adult), autos (one new car per year), radio (6 hours per day)
- Typically the key estimates are not that sensitive to plausible market definitions.
- However, market size sometimes affects numerical performance with richer models.

[1] This is only partly true in the sense that we still have to define the set of products we are interested in (e.g., in scanner data are milk and beer in the product set?).

Multinomial Logit Model

Estimation with Aggregate Data

- An obvious concern is price endogeneity because prices are correlated with ξ.
  - x (obs. characteristics) may also be correlated with ξ, but people tend to ignore this problem.
- Dealing with endogeneity in a non-linear setting is hard, therefore we'd like to get a **linear estimating equation**.

Multinomial Logit Model

Estimation with Aggregate Data: Inversion

- Berry (1994), following Hotz and Miller, proposed inverting out market shares to get mean utilities
$$s = G(\delta), \quad \delta = G^{-1}(s) = x\beta - \alpha p + \xi$$
i.e., solving a system of K equations in each market for K unknowns (δ_K). Here we're assuming we observe market shares.

- For the logit, there is a beautiful, simple form of inversion
$$\ln s_{kt} - \ln s_{0t} = \delta_{kt} - \delta_0 = x_{kt}\beta - \alpha p_{kt} + \xi_{kt}$$
(you'll agree it's beautiful when you work with other models!)

Multinomial Logit Model

Estimation with Aggregate Data: 2SLS

- We can then estimate using 2SLS

$$\underbrace{\ln s_{kt} - \ln s_{0t}}_{y} = \delta_{kt} - \delta_0 = x_{kt}\beta - \alpha p_{kt} + \xi_{kt}$$

- Use 2SLS because we expect p to be endogeneous (correlated with ξ)
  - requires instruments: e.g., cost shifters or Hausman instruments
  - or 'BLP' instruments to be discussed in more detail below
- As with homogenous products expect demand to be more elastic when control for endogeneity.

- Suppose we observe a panel of 1,000 people from each of 20 different markets.
  - E.g., individual household panel where prices differ across markets, but the set of products is the same
  - In each period we observe the price of each product in the choice set (for example, this might come from accompanying aggregate data)
- How to proceed?

Multinomial Logit Model

Estimation with Individual Data

Estimate a multinomial choice model using ML in a single-step

- In STATA you would use

```
clogit choice price x1 x2 ..., group(panelist_wkid)
```

  - conditional logit

- In MATLAB you would define the log-likelihood yourself and use an optimization tool.
  - As we shall see, doing so requires you to be able to choose some sensible options and interpret MATLAB output.
  - However, because the log-likelihood for this model is globally concave almost any choice is going to work well
    - but some choices will work quicker than others

Some problems have a very specific structure that should be exploited, e.g., constrained least squares

`lsqlin`

- This can be useful in some settings: e.g., when all you need to find is a set of weights on a given set of types.
- Algorithms used are guaranteed to converge, find a unique solution and be quick.

Detour: Optimization

In MATLAB's Optimization Toolbox, What Are the Basic Options?

`fminsearch`

- Uses non-derivative based method known as Nelder-Mead (also amoeba).
- Your best, although still problematic, bet when your objective function has flat spots.
  - This is common where simulation of discrete outcomes involved.
  - Often used in the context of entry models.
- Think of it as a simplex trying to roll downhill like a ball.
- Often ends up evaluating the objective function many, many times.

Detour: Optimization

In MATLAB's Optimization Toolbox, What Are the Basic Options?

```
options=optimset('Display','iter','TolX',1e-8,...
    'TolFun',1e-8,'MaxIter',1e5,'MaxFunEvals',1e6);
[est,fval,exitflag]=fminsearch('mylogitmlfun',...
    start,options,otherx1,otherx2,...);
```

`exitflag>=1` is good (evidence of a converged solution)

`exitflag<=0` is not good, means you need to try again or at least look carefully ...

Detour: Optimization

In MATLAB's Optimization Toolbox, What Are the Basic Options?

`fminunc`

- Derivative-based methods
  - Which one depends on the options and information that you provide.
  - Uses first-derivative and possibly second-derivative information to move from a current guess to a hopefully better guess.

Detour: Optimization

In MATLAB's Optimization Toolbox, What Are the Basic Options?

```
options=optimset('Display','iter','TolX',...
    1e-8,'TolFun',1e-8,...
    'MaxIter',1e5,'MaxFunEvals',1e6,'GradObj',...
    'on','Hessian','off');
[est,fval,exitflag]=fminunc('mylogitml fun',...
    start,options,otherx1,otherx2,...);
```

Detour: Optimization

In MATLAB's Optimization Toolbox, What Are the Basic Options?

- You should always choose `TolX` and `TolFun` to be 1e-6 or smaller when coefficients and variables are scaled in the range say [-10,10]
  - For final estimates, try to use 1e-8 or 1e-10
- Providing derivatives can increase performance significantly.
  - Programming analytic derivatives can be a pain (potentially reduced using Mathematica or Maple).
  - But it is usually worth the effort once you have chosen your model unless the objective function is very well-behaved.
  - Whenever you program derivatives you should check them.
- Providing Hessians can also increase performance significantly
  - but it can be exceptionally difficult outside basic logit.
  - but it is worth considering.

Most algorithms are based on **Newton's Method**

- A good method because it converges quadratically to a minimum, at least in its neighborhood.
- Idea: suppose you have a one dimensional problem, a **quadratic** function f and you are looking for a point x^* where $f(x^*) = 0$.
    - If you are currently at a point x_0, you would find x^* as

$$x^* = x_0 - \frac{f(x_0)}{f'(x_0)}$$

- In our problem we want $f'(x^*) = 0$ so we would use

$$x^* = x_0 - \frac{f'(x_0)}{f''(x_0)}$$

Detour: Optimization

Why Are Derivatives Worthwhile?

- The hope is that a quadratic approximation to our function is at least going to take us to a better location than x_0.
- So derivatives are central to how we move, and accurate, quickly computed derivatives will be very helpful.
- Also in quasi-Newton procedures (e.g., BFGS) gradients are used to approximate the Hessian, so accuracy will pay off in both the numerator and denominator.

Detour: Optimization

Why Are Derivatives Worthwhile?

If you can't give analytic derivatives, MATLAB's alternative is to use numerical **finite differences**

$$\frac{\partial f}{\partial x_i} = \frac{f(x_1, \ldots, x_i + h, \ldots, x_C) - f(x_1, \ldots, x_i, \ldots, x_C)}{h}$$

- So you have to compute the objective function at least # of parameters + 1 times.
- In practice these one-sided derivatives may not be that accurate (or, it is hard to choose *h* correctly).
- Two-sided derivatives are more expensive (2*# of parameters), but more accurate.
- Inaccurate derivatives can cause the algorithm to have to check many more points (in a line search), before choosing the next iteration.

Detour: Optimization

Writing the Function

Suppose you are writing a function that calculates the objective function and the analytic derivatives. Often the derivative calculations will take several times the time that it takes to calculate the objecvtive function. Therefore, it is a good idea to only calculate the derivatives when MATLAB actually wants the derivatives.

```
function [f,g]=mylogitml_fun(parameters,otherx1,otherx2)
%calculate the objective function
f=...
if nargout>1,
g=...
end
```

Detour: Optimization
How to Check Derivatives?

Suppose your function returns the objective function and the gradient, then ...

```
[f,g]=mylogitml_fun(guess,otherx1,...)
numg=[];dh=1e-7;
for i=1:length(guess),
altguess=guess;altguess(i)=altguess(i)+dh;
altf=mylogitml_fun(altguess,otherx1,...); numg=[numg;(altf-f)./dh];
end
```

Detour: Optimization

- When you do this you may find that `g` and `numg` are different for some parameters.
  - How much will likely vary with the scale of the coefficient and the variable.

- This may reflect the fact that the `dh` you want to use is not the same for different parameters.

- If MATLAB is going to use numerical derivatives, then you can control the dh that it uses using

  `DiffMinChange,DiffMaxChange`

  options, when setting `optimset`.

Detour: Optimization Hessians

- Ideally you would provide a Hessian (second derivative matrix)
- This is pretty easy for a multinomial logit
  - But also somewhat unnecessary because the multinomial logit has a very nice likelihood function!
  - You should check your calculation, following the same logic as above.
- If not, MATLAB might either try to calculate a Hessian at each iteration (by taking derivatives of gradients) or can use gradients to form estimates of the Hessian.
  - When it is time consuming to calculate derivatives, this approach is best.

```
options=optimset('Display','iter','GradObj','on','TolX',1e-10,
    'MaxFunEvals',1e5,'MaxIter',1e4,'LargeScale','off',...
    'HessUpdate','bfgs');
```

Detour: Optimization

A Useful Hint

When optimization performs badly (in either STATA or MATLAB) there is often a simple cause and a straightforward fix.

It is very important that the Hessian is well-conditioned, and there are at least two situations where this will not be true:

1. Xs are perfectly collinear or almost collinear
    - In linear regression STATA will detect this and drop variables appropriately.
    - In non-linear estimation its algorithm for doing this is not perfect especially with a big dataset.
    - In MATLAB it is always a good idea to check with `rcond`.
        - A low number (e.g., $< 1e - 16$) indicates a significant problem.
    - Be ready to drop or combine variables.

Detour: Optimization

A Useful Hint

2. The scale of the parameters or the Xs is very different (e.g., 50, 000 and $5e - 3$).

- In this case be ready to redefine units, which is simple.
- It is useful to scale all of the X variables (apart from the constant!) so that they have mean zero and a std of 1.
- Of course, you may want to reverse the scaling to report the results.

Multinomial Logit Model
Estimation with Individual Data

However, when we use

```
clogit choice price x1 x2 ..., group(panelist_wkid)
```

we have not quite estimated the model we had above, i.e., we have estimated a model where

$$U_{ijt} = x_{jt}\beta - \alpha p_{jt} + \varepsilon_{ijt}$$

not

$$U_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

i.e., we have not allowed for the ξ_{jt} term that would tend to make choices correlated across individuals (conditional on observables).

- In doing so we have also removed (by assumption) the source of the endogeneity problem.

Multinomial Logit Model

Estimation with Individual Data

How you feel about this/address it depends on the context:

- Marketers will often argue that price reductions in scanner data are exogenous to period/product-specific demand shocks.
- In this case you should assume $\xi_{jt} = \xi_j$ and include product fixed effects to deal with the fact there is a time-invarient product effect.
- As long as you have enough panelists and weeks of data this could work (remember we have non-linear model).[2]
- The βs for product characteristics could then be recovered in a second-stage projection.
- Really depends on whether you can actually control for everything else consumers see and which might be relevant for the pricing decision.

[2] With enough data panellists you could estimate a product-market-week fixed effects and then estimate price coefficients in a linear second stage. But this essentially amounts to claiming that you have enough data to see aggregate market shares.

- The issue is whether you like the assumption of exogenous price variation?
- When do we see sales of Coke, Pepsi and potato chips?

Multinomial Logit Model

The Problem

- Implies restrictive and very unreasonable substitution patterns

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

- Why are these unreasonable? Own price elasticities ...
  - *conditional on market share*, higher priced goods have more elastic demand
    - Given single-product profit maximization, cheaper goods will have higher mark-ups (BMW, Kia).
    - In some settings the market shares of all products will be fairly small and similar.
  - *conditional on price*, elasticity decreases with market share
    - products with bigger shares will have bigger mark-ups
- Cross-price elasticities are just functions of prices and shares of other products, not their characteristics

Multinomial Logit Model

IID and IIA

- The cross-price and own-price properties reflect the IID nature of the ε shocks across products.
  - In the model these are the only reason why all consumers don't buy the same thing.
  - Not just a function of logit.
- IID is not the same as the famous **Independence of Irrelevant Alternatives (IIA)** property

$$\text{IIA} \Leftrightarrow \frac{s_j}{s_k} \text{ does not depend on the presence or prices of other products.}$$

  - The logit model has the IIA property, but other distributions (e.g., normal) do not.
  - The red bus/blue bus problem (is it a problem?).

Welfare in Discrete Choice Logit Models

Suppose indirect utility is

$$x_{jt}\beta_i - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

with extreme value εs. Then expected utility/welfare/inclusive value per person is

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

- If price is linear, divide by α_i to get dollar effect.
- Analysis is more complicated if we allow for income effects or some other non-linearity (e.g., demand depends on log price).

Welfare in Discrete Choice Logit Models

A convenient result is that in a pure logit model

$$\omega_{iAt} = \ln\left(\sum_{j \in A} \exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})\right) = \ln\left(\frac{1}{s_{0t}}\right)$$

- i.e., if we know the market share of the outside good then we know utility.
- Now let's introduce the red bus/blue bus problem.

Red Bus/Blue Bus Problem

This famous example illustrates the dangers associated with the IIA, or at least an unthinking application of the model.

Suppose we are interested in commuting mode choice.

- Our data: equal numbers of people go by car, ride the bus, which is red, and walk (the outside option).
- Implied δs.

Mode	**Share**	**Implied δ**
Car	0.33	0
Red Bus	0.33	0
Walk	0.33	0 (assumed)

- Welfare: $\ln(1/0.33) = \ln(3)$

Red Bus/Blue Bus Problem

Now suppose we add a blue bus to the choice set, which is otherwise identical to the red bus.

- It is natural to assume that $\delta_{blue} = 0$, and that the proportion of people taking the red and blue buses will be the same.
- But, under IIA, the proportion of people choosing the red bus and walk should still be equal.
- i.e., we will predict

Mode	**Share**	**Implied δ**
Car	0.25	0
Red Bus	0.25	0
Blue Bus	0.25	0
Walk	0.25	0 (assumed)

and welfare will be $\ln(4)$.

Red Bus/Blue Bus Problem

- What doesn't seem right here is that we are mispredicting market shares, by failing to recognize that the red bus and blue bus are identical i.e., what likely happens is

Mode	**Share**
Car	0.33
Red Bus	0.167
Blue Bus	0.167
Walk	0.33

  in which case we would get the same welfare as in the first example ($\ln(3)$).

- So the issue is that under IIA we will mispredict market shares when we add products, because we ignore that the new product may be particularly similar to existing products, and therefore get misleading welfare results.

Nested Logit

- The nested logit model builds in correlation among the residuals for different products based on an a priori discrete classification by the researcher.
  - If we were using normally distributed εs we'd do this by allowing for correlation of the εs.
  - But estimating multinomial probit models is hard once we go beyond 3 or 4 options (in STATA you would use `mprobit`).
- β and α still same across consumers, but now a consumer can have a taste for a group of products.
- It requires the researcher to pre-specify groups of products (nests) prior to estimation.
  - It usually only makes sense to group on **discrete** characteristics.
  - e.g., radio formats, types of auto (SUV, minivan, sports), types of beer.
  - Nesting can have multiple levels.

Nested Logit

- For a one-level nested logit model

$$
\begin{aligned}
U_{ijt} &= x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \zeta_{ig(j)t} + (1-\sigma)\varepsilon_{ijt} \\
&= \delta_{jt} + \zeta_{ig(j)t} + (1-\sigma)\varepsilon_{ijt}
\end{aligned}
$$

  ε is iid Type I EV across products (as before), $\zeta_{ig(j)t}$ is a group specific taste of the consumer, σ measures the relative weight of idiosyncratic and group preferences

$$
v_{ijt} = \zeta_{ig(j)t} + (1-\sigma)\varepsilon_{ijt}
$$

- There is a unique distribution of ζ such that v_{ijt} is also extreme value (Cardell (1999)), and which provides a nice and invertable form of the market share integral.

- Note: this presentation follows Berry (1994); in other presentations it is often assumed that the σs differ across groups at the same level, and that one is working with $\tau = 1 - \sigma$.

Nested Logit

Market shares (choice probabilities):

$$\begin{aligned} s_{jt} = s_{j|g(j)} s_{g(j)} &= \frac{e^{\frac{\delta_j}{1-\sigma}}}{D_g} \frac{D_{g(j)}^{1-\sigma}}{\left[\sum_g D_g^{1-\sigma}\right]} \\ &= \frac{e^{\frac{\delta_j}{1-\sigma}}}{D_g^{\sigma} \left[\sum_g D_g^{1-\sigma}\right]} \end{aligned}$$

where $D_g = \sum_{j \in g} e^{\frac{\delta_j}{1-\sigma}}$

Nested Logit

- Substitution patterns:
  - now depend on nest shares and whether products are in the same nest

$$\text{MNL: } \eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = -\alpha p_{jt}(1 - s_{jt}) \text{ if } j = k$$

$$\alpha p_{kt} s_{kt} \quad \text{otherwise}$$

$$\text{NL: } \eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = -\alpha p_{jt} \left(\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} s_{jt|g(j)} - s_{jt} \right) \text{ if } j = k$$

$$\alpha p_{kt} s_{kt} \quad \text{if } j \neq k \quad \text{and they are in different nests}$$

$$\alpha \frac{s_{kt}}{s_{jt}} p_{kt} \left(\frac{\sigma}{1-\sigma} s_{jt|g(j)} + s_{jt} \right) \text{ if } j \neq k \quad \text{and they are in the same nests}$$

Nested Logit

- Within-nest cross-price elasticities just depend on market shares and prices, but will be greater than cross-nest elasticities.
  - People who like light beer more likely to switch to other light beers.
- As $\sigma \to 1$, substitution is only within nest.
- If $\sigma = 0$, back to the multinomial logit model.
- If σ is outside [0,1] the model is not consistent with utility maximization.

Nested Logit

Estimation: Individual Data

- The same conceptual issues we discussed with the logit model still apply (don't include ξ, unless have enough data for product fixed effects).
- In STATA, can use the `nlogit` command
  - but it will try to allow for different nesting coefficients (our σs) on each nest, and you need to use constraints to prevent this.
  - it can also be slow for large datasets.
  - can be done sequentially (from the bottom of the tree upwards) using "inclusive values", with a loss of efficiency.
  - most well-known example and reference is Goldberg (1995, ECMA), but the STATA manual is also good!

Nested Logit

Estimation: Aggregate Data

- With aggregate data, Berry (1994)

$$\begin{aligned}\ln s_{kt} - \ln s_{0t} &= \sigma \ln s_{k,g(k),t} + \delta_{kt} - \delta_0 \\ &= \sigma \ln s_{k,g(k),t} + x_{kt}\beta - \alpha p_{kt} + \xi_{kt}\end{aligned}$$

  where $s_{k,g(k),t}$ is k's share of its nest; i.e., we still have an easy inversion.

- Estimation using 2SLS:
  - need to instrument for price as before
  - and must instrument for $\ln s_{k,g(k),t}$ which will definitely be endogenous
  - i.e., product fixed effect arguments can no longer avoid endogeneity

Nested Logit

Extensions

- Multiple level of nests (e.g., Goldberg 1995).
- Overlapping nests (Bresnahan, Stern and Trajtenberg (1997)).
  - Essentially allows you not to commit to the nesting sequence while retaining some of the tractability of the NL specification.
  - Arcidiacono, Ridley et al. (2012) use in the context of pharma where may be difficult to order nests for brand/generic, drug form, drug type.
- Limitations:
  - Have to impose some type of discrete nesting a priori.
    - It is natural in some examples, e.g., radio formats.
  - May be arbitrary and key characteristics may be continuous.
  - α, β still constant.

Random Coefficient Logit Model (Mixed Logit)

- Generalizes the logit model to provide a more flexible representation of heterogeneous preferences than the nested logit.
    - Much easier to incorporate heterogeneity on continuous characteristics.
- It is often called the "BLP model" after the most famous application.
- Utility, e.g.,

$$U_{ijt} = x_{jt}\beta_i - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

$$\text{Where } \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \prod D_i + \sum v_i$$

$$v_i \sim P_v(v), D_i \sim \widehat{P_D(D)}$$

D_i are observed demographics (age, sex, race) and $P_v(v)$ are known parametric distributions.

Random Coefficient Logit Model (Mixed Logit)

- In practice one might decide that price coefficient is log-normal, and that taste for sugar content is normal.
- The parameters to estimate are α, β, $\prod$, $\sum$ (or other appropriate distributional parameters).
- Handles heterogeneous tastes for discrete or continuous characteristics.

Random Coefficient Logit Model

Do We Need It?

- If you have individual data then you could have allowed for individual-characteristic x product characteristic interactions in a multinomial logit model.
  - black x urban radio, have kids x yogurt drinks
  - experience x price (as a proxy for search ability)

- the RC model is different in:
  1. With aggregate data it links purchases to individual characteristics, not just market averages (e.g., % black).
  2. With either individual or aggregate data it allows for a random component of preferences that is not correlated with demographics (v_i).

Random Coefficient Logit Model

Do We Need It?

- Intuitively the v_is are attractive.
  - Clearly there is significant unobserved variation in tastes for characteristics as well as products.
  - This could matter for predictions as well as welfare.
- But in practice:
  - They often turn out to be small, statistically insignificant or sensitive.
    - Probably reflects lack of identifying variation in most datasets.
  - It can make estimation significantly harder, requiring simulation even with individual data.

Random Coefficient Logit Model

Interpretation

- It's useful to remember that the random coefficient logit model is a "person-by-person" logit model.
- Each person or type of person has multinomial logit preferences (and hence expected substitution patterns associated with the simple logit model).
- But in aggregate the elasticities are more flexible because we integrate out over a population whose preferences (and hence market shares) and elasticities are different.
  - Why? light beers will tend to be bought by people who have very strong preferences for light beer.

Random Coefficient Logit Model

Interpretation

- It is also useful to define:

$$U_{ijt} = \delta_{jt} + x_{jt}(\beta_i - \beta) - (\alpha_i - \alpha)p_{jt} + \varepsilon_{ijt}$$
$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

  i.e., **"mean utility"** now means the average utility of the **baseline-type of consumer**, not the population average.

- Call (β, α) the linear parameters.

- Non-linear parameters: $(\prod, \sum)$.

Random Coefficient Logit Model

Market Shares/Choice Probabilities

- Market shares are given by the following integration:

$$s_{jt} = \int \dots \int_{A_j} f(\beta, \alpha, \varepsilon)\, d\beta\, d\alpha\, d\varepsilon$$

  which is hard to compute analytically.

- The standard approach is to use Monte Carlo (simulation) integration, e.g.,

$$s_{jt} = \frac{1}{R} \sum_{r=1}^{R} s_{rjt}(\beta_r, \alpha_r)$$

  where r represents a draw from the distribution of β and α.

- This exploits the fact that we have an analytic formula for the market shares of any individual, because they are just a logit consumer.

Detour: Numerical Integration

To solve or estimate a model, or calculate welfare, we will often need to integrate; examples:

$$s_{kt} \equiv \Pr(\text{choose } k) = \int \ldots \int_{A_j} f(\varepsilon_1, \ldots, \varepsilon_J \mid \delta)\, d\varepsilon_1 \ldots d\varepsilon_J$$

where A_k defines the region of ε space where product k gives highest utility

$$s_{kt} \equiv \Pr(\text{choose } k) = \int_{\underline{\beta_1}}^{\overline{\beta_1}} \ldots \int_{\underline{\beta_C}}^{\overline{\beta_C}} s_{ikt}(\overrightarrow{\beta_i}) g(\beta_1, \ldots, \beta_C)\, d\beta_1 \ldots d\beta_C$$

Detour: Numerical Integration

- In an ideal world we'd always be able to compute analytic integrals.
- In some settings analytic formula **are** available.

Example

Multinomial logit demand with $u_{ik} = \delta_k + \varepsilon_{ik}$

$$\Pr(i \text{ chooses } k) = s_k = \frac{\exp(\delta_k)}{\sum_j \exp(\delta_j)}$$

Example

Multinomial logit demand with $u_{ik} = \delta_k + \varepsilon_{ik}$

$$\text{Expected utility from choosing best option(EU)} = \log\left(\sum_j \exp(\delta_j)\right)$$

- Why logit so useful (cf. normal).

Detour: Numerical Integration

- For low dimensional integration, we may be able to use some form of **quadrature**.
- Which is a way of computing integrals using squares, rectangles and cubes.
- There are different methods for choosing the width of the pieces (e.g., Gauss-Hermite, Simpson's method).
- These can be very accurate for one or two dimensions, but they suffer from the curse of dimensionality.
- If we need to calculate derivatives small inaccuracies may matter and it may be slow.

Detour: Numerical Integration

Examples with a Normal Distribution

- Let's suppose we were determined to use a normal distribution for our preference residuals.
  - This might be one route to modeling correlations.
- How could we calculate $\Pr(i \text{ chooses } k)$ and EU?
- The answer will depend on:
  - The size of the choice set (number of dimensions of integration).
  - The required degree of accuracy.
- With two or three choices, one could try to use a procedure like Gauss-Hermite quadrature.
- But more generally we will often use **simulation**.

Detour: Numerical Integration

Naive Frequency Monte Carlo Simulator

The simplest simulator of an expectation of a function will:

- Take draws from the density.
- Calculate the function for each draw.
- Take the average!

Example

Pr(i chooses k) would be computed

$$\sum_{s=1}^{S} \frac{I(\text{choice } k \text{ maximizes utility in sim. } s)}{S}$$

where we take random draws from a normal distribution.

Suppose 4 choices, $u_{ik} = \delta_k + \varepsilon_{ik}, \delta_{0,\ldots,3} = 0, -2, 1, 2$ and $\overrightarrow{\varepsilon_i}$ is distributed iid normal, mean 0, variance 1.

How many draws are required to get the choice probabilities to converge?

Detour: Numerical Integration

Naive Frequency Monte Carlo Simulator: Performance

Suppose 4 choices, $u_{ik} = \delta_k + \varepsilon_{ik}, \delta_{0,\ldots,3} = 0, -2, 1, 2$ and $\overrightarrow{\varepsilon_i}$ is distributed iid normal, mean 0, variance 1. .

How many draws are required to get the choice probabilities to converge?

sims	Pr_0	Pr_1	Pr_2	Pr_3	EU
5	0	0	0.2	0.8	2.8596
20	0.05	0	0.2	0.75	2.2802
100	0.04	0	0.28	0.68	2.1280
1,000	0.55	0	0.237	0.708	2.2208
10,000	0.0466	0.0006	0.2161	0.7367	2.2180
100,000	0.0478	0.0004	0.2223	0.7294	2.2242
truth	0.0471	0.0005	0.2240	0.7285	2.2249

Detour: Numerical Integration

Improved Simulators

- The results above indicate that you need a lot of draws to get accurate probabilities.
- While the frequency simulator is unbiased for the probability it will be biased for the log(probability).
  - and an SML estimator will only be consistent as $S \to \infty$ even if $N \to \infty$
- Can we design a better simulator?
  - By taking draws in a way that reduce the variance of our estimator.

Detour: Numerical Integration

Antithetic Sampling

- When sampling antithetically, for each draw $\{\varepsilon_0, ..., \varepsilon_3\}$ we also use $\{-\varepsilon_0, ..., -\varepsilon_3\}$.

sims	Pr_0	Pr_1	Pr_2	Pr_3	EU
5	0.2	0	0	0.8	2.7014
20	0.1	0	0.25	0.65	2.3988
100	0.06	0.01	0.23	0.70	2.2437
1,000	0.0450	0.001	0.215	0.7390	2.2124
10,000	0.0476	0.0006	0.2158	0.7360	2.2227
100,000	0.0471	0.0005	0.2227	0.7298	2.2256
truth	0.0471	0.0005	0.2240	0.7285	2.2249

- This is still a Monte Carlo simulator because we are taking random draws.

- We can also avoid taking random draws and instead use some other type of sampling procedure.

- An example is **Halton draws**.
  - a halton sequence is based on a prime number base and gives draws on an interval (0,1).
  - for example, for base 2: $\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, \ldots$
  - for base 3: $\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \ldots$
  - so we could form a 2-d sequence $(\frac{1}{2}, \frac{1}{3}), (\frac{1}{4}, \frac{2}{3}), \ldots$

- To use these in our example, we would invert back from the cdf to get the draws of the εs.

- When one goes to high dimensions one can do better by using *scrambled* Halton sequences.

Applied to our example,

sims	Pr_0	Pr_1	Pr_2	Pr_3	EU
5	0.2	0	0.2	0.6	2.3463
20	0.05	0	0.3	0.65	2.1590
100	0.05	0	0.21	0.74	2.1734
1,000	0.0480	0	0.222	0.73	2.2149
10,000	0.0472	0.0005	0.2238	0.7285	2.2233
100,000	0.0471	0.0005	0.2239	0.7285	2.2247
truth	0.0471	0.0005	0.2240	0.7285	2.2249

Further examples, e.g., Sobol draws

sims	Pr_0	Pr_1	Pr_2	Pr_3	EU
5	0	0	0	1	2.6294
20	0	0	0.25	0.75	2.2826
100	0.04	0	0.20	0.76	2.2452
1,000	0.045	0	0.218	0.737	2.2247
10,000	0.0478	0.0005	0.2244	0.7273	2.2246
100,000	0.0472	0.0004	0.2239	0.7284	2.2249
truth	0.0471	0.0005	0.2240	0.7285	2.2249

Detour: Numerical Integration

Comparison Table for 10,000 draws

sims	Pr_0	Pr_1	Pr_2	Pr_3	EU
truth	0.0471	0.0005	0.2240	0.7285	2.2249
naive	0.0466	0.0006	0.2161	0.7367	2.2180
antithetic	0.0476	0.0006	0.2158	0.7360	2.2270
halton	0.0472	0.0005	0.2238	0.7285	2.2233
sobol	0.0478	0.0005	0.2244	0.7273	2.2246

Detour: Numerical Integration

Random Coefficient Logit setting

In the RC-logit setting we are helped by the fact that for a given set of coefficients we have analytic formulae for computing shares/choice probability

- Therefore **we should not simulate the logit errors to find individual choices**.
- Instead, only simulate the random coefficients and, for each coefficient, calculate the choice probabilities analytically.
- At the very least this will definitely make sure we avoid the log(0) problem!

- It is important to be able to approximate the integral accurately.
- As we will see some applications that use a small number of randomly drawn points are not that accurate.
- Better implementations use a large number of Halton draws.
- Of course, initial estimates may use a smaller number of draws than final ones.

Random Coefficient Logit Model

Elasticities

The price elasticities of the market shares, s_{jt} are

$$\eta_{jkt} = \frac{\partial s_{jt} p_{kt}}{\partial p_{kt} s_{jt}} = \begin{cases} -\dfrac{p_{jt}}{s_{jt}} \int \alpha_i s_{ijt}(1 - s_{ijt})\, d\hat{P}_D^*(D)\, dP_v^*(\upsilon) & \text{if } j = k \\ \dfrac{p_{kt}}{s_{jt}} \int \alpha_i s_{ijt} s_{ikt}\, d\hat{P}_D^*(D)\, dP_v^*(\upsilon) & \text{otherwise} \end{cases}$$

where $s_{ijt} = \exp(\delta_{jt} + \mu_{ijt})/[1 + \sum_{k=1}^{K} \exp(\delta_{kt} + \mu_{ikt})]$ is the probability of individual i purchasing product j. Now the own-price elasticity will not necessarily be driven by the functional form. The partial derivative of the market shares will no longer be determined by a single parameter, α. Instead, each individual will have a different price sensitivity, which will be averaged to a mean price sensitivity using the individual specific probabilities of purchase as weights.

- Obviously the accuracy of these elasticities will depend on the same factors as the accuracy of the market share calculations.

Random Coefficient Logit Model

Estimation with Individual Data

- There are some examples of papers estimating RC models with individual data:
  - E.g., Leslie, Sorensen and Chu, "Bundle-Size Pricing as an Approximation to Mixed Bundling", AER, 2011.
- You can estimate this model using STATA using a user-created program mixlogit.
- To install it,

```
ssc install mixlogit
```

- To calculate the integral it makes use of simulation using Halton draws.

Random Coefficient Logit Model

Recall...

- Generalizes the logit model to provide a more flexible representation of heterogeneous preferences.

- Utility, e.g.,

$$U_{ijt} = x_{jt}\beta_i - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

$$\text{Where } \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \prod D_i + \sum v_i$$

$$v_i \sim P_v(v), D_i \sim \widehat{P_D(D)}$$

- The parameters to estimate are α, β, $\prod$, $\sum$ (or other appropriate distributional parameters).
  - Usefully define $\prod$, $\sum$ as non-linear parameters.
  - α, β are linear.

Estimation Using Aggregate Data Following BLP/Nevo

- The random coefficient model is frequently estimated using aggregate data even though the model has heterogeneity in individual preferences.
  - The idea is that the RC model will allow us to fit observed substitution patterns that contradict MNL assumptions.

- Key is to form a linear estimating equation, based on the δs just like we did in estimating the multinomial logit and nested logit models.

Estimation Using Aggregate Data Following BLP/Nevo

- In the logit case

$$\xi_{jt}(\delta,\beta,\alpha) = \underbrace{\delta_{jt}}_{\equiv \ln s_{jt} - \ln s_{0t}} - x_{jt}\beta + \alpha p_{jt}$$

  and we assume that ξ_{jt} is uncorrelated with x_{jt} and cost shifters (or other instruments for price). Gives moment conditions.

$$E(\xi_{jt}(\delta,\beta,\alpha) \mid x_{jt}, Z_{jt}) = 0$$

- Which can be implemented using GMM or (more simply) 2SLS.

- The same idea applies here, except that.
  - $\delta_{jt} \neq \ln s_{jt} - \ln s_{0t}$ (the inversion is more complicated, recall that now δ is the mean utility of baseline consumers).
  - We have additional non-linear parameters reflecting heterogeneity of preferences.

Estimation Following BLP/Nevo

Inversion

Therefore a key step will be the inversion from observed market shares to the δs.

- The multinomial logit/nested logit cases this was simple

$$\delta_{jt} \equiv \ln s_{jt} - \ln s_{0t}$$

$$\delta_{jt} \equiv \ln s_{jt} - \ln s_{0t} - \sigma \ln s_{jg(j)}$$

- Note there are two properties:
  1. The mapping from δs to market shares is unique in both directions.
  2. The inversion is computationally simple.

Estimation Following BLP/Nevo

Inversion

For the RC model, Berry/BLP prove:

1. Uniqueness (for given non-linear parameters $\prod$, $\sum$); and,
2. The set of equations form a **contraction**.

- This means that an iterative procedure can be used to solve for the δs that is **guaranteed** to work.

- We can iterate

$$\delta_{.t}^{h+1} = \delta_{.t}^{h} + \ln S_{.t} - \ln\left(\widehat{S_{.t}}(\delta_{.t}^{h}, p_{.t}, x_{.t}, P \mid \prod, \sum)\right)$$

  where the $.t$ notation denotes a vector of values for market or time t, P represents distributional assumptions (e.g., observed income distribution) and $\widehat{S_{.t}}$ are values predicted by the demand model.

- We stop iterating when $\max(\text{abs}(\delta_{.t}^{h+1} - \delta_{.t}^{h}))$ is less than some tolerance, which we will see needs to be very demanding e.g., $1e-14$.

- Given values of $\hat{\delta}$ (calculated from the contraction mapping, given values of the non-linear parameters), we can calculate moments

$$\xi_{jt}\left(\hat{\delta}\left(S, p_{.t}, x_{.t}, P \mid \prod, \sum\right), \beta, \alpha\right) = \hat{\delta}\left(S, p_{.t}, x_{.t}, P \mid \prod, \sum\right) - x_{jt}\beta + \alpha p_{jt}$$

$$E\left(\xi_{jt}\left(\prod, \sum, \beta, \alpha\right) \mid x_{jt}, Z_{jt}\right) = 0$$

  where **we need enough instruments to identify the non-linear, as well as the endogenous linear, parameters**.

- As we will see in a moment, the linear parameters factor out so we only need to do a numerical search/optimization over the non-linear parameters.

- Let's discuss the details of estimation and then what additional instruments we can use.

Random Coefficient Logit Model

Estimation - Cookbook: What You Need

Assume we have:

- A set of data on the same products from different markets.
  - E.g., data on sales (market shares given definition of market), prices of 25 top cereal brands in 100 markets.
- We observe some characteristics of each product.
  - Sugar content, whether it goes soggy in milk.
- Distribution of demographics in each market: e.g., joint distribution of income, kids, age, education (i.e., typically think of households).
- Set of instruments for prices (discuss in more detail below)

Random Coefficient Logit Model

Estimation - Pre-Steps

- For each market, draw *NS* households from the market-specific distribution of demographic characteristics.
  - E.g., household 1: income $85k, 3 kids, hoh: 50 years old, college-educated.
- For each simulated household: draw v_is for each characteristic (let's assume these are normal).
- Note: all of these draws **must** stay constant during the estimation procedure.
- Specify the linear and non-linear parameters.
  - Choose starting values for the non-linear parameters.

Main Estimation Loop

Preliminaries: (i) create a simulated population size *ns* with demos D and draws v, (ii) pick starting parameters $\prod^0$, $\sum^0$

$\downarrow$

Compute $\delta^k(S, \prod^k, \sum^k)$ using contraction mapping procedure

$\downarrow$

Calculate α^k, β^k analytically and compute ξ^k

$\downarrow$

Calculate objective function and maybe gradient (can be analytic)

$\downarrow$

Meet stopping criteria?

No → Update $\prod^k$, $\sum^k$; $k = k + 1$ (return to Compute δ^k)

Yes → Done

Contraction Mapping Procedure

Pick δ^0

$\downarrow$

Calculate $S(\delta^w, \prod^k, \sum^k)$ using the simulated population

$\downarrow$

Is $\text{abs}(S(\delta^w, \prod^k, \sum^k) - S^{\text{DATA}}) < \text{tol}^{\text{CM}}$?

Yes → Return δ^w

No → $\delta^{w+1} = \delta^w + \ln(S) - \ln(S(\delta^w, \prod, \sum))$ (return to Calculate S)

Random Coefficient Logit Model

Estimation - The Routine

- The GMM objective function is

$$\min_{\theta} \hat{\xi}(\hat{\theta})' W \Phi^{-1} W' \hat{\xi}(\hat{\theta})$$

where W is a matrix of the instruments and exogenous product characteristics and $\hat{\theta}$ represents all of the parameters; Φ^{-1} is a weighting matrix (best is inverse covariance matrix of the moments)

$$\hat{\xi}(\hat{\theta}) = \hat{\delta}(S, p_{.t}, x_{.t}, P \mid \widehat{\prod}, \widehat{\sum}) - x_{jt}\hat{\beta} + \hat{\alpha} p_{jt}$$

Note that for given values of the mean utilities, optimal $\hat{\alpha}$ and $\hat{\beta}$ (linear parameters) can be solved for analytically. If $X_1 = [p\ X], \theta_1 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$$\hat{\theta}_1 = (X_1' W \Phi^{-1} W' X_1)^{-1} X_1' W \Phi^{-1} W' \hat{\delta}(S, p_{.t}, x_{.t}, P \mid \widehat{\prod}, \widehat{\sum})$$

- So now only have to search over $\widehat{\prod}$, $\widehat{\sum}$

- The search over $\prod$, $\sum$ is numerical (i.e., no analytic solution).

- For a given guess of $\widehat{\prod}$, $\widehat{\sum}$ solve for $\hat{\delta}(S, p_{.t}, x_{.t}, P \mid \widehat{\prod}, \widehat{\sum})$ using the contraction mapping
$$\hat{\delta}_{.t}^{h+1} = \hat{\delta}_{.t}^{h} + \underbrace{\ln S_{.t}}_{DATA} - \ln\Big(\widehat{S_{.t}}(\delta_{.t}, p_{.t}, x_{.t}, P \mid \prod, \sum)\Big)$$
Implementation requires the calculation of $\widehat{S_{.t}}(\delta_{.t}, p_{.t}, x_{.t}, P \mid \prod, \sum)$. Use our simulated consumers!

Random Coefficient Logit Model

Estimation - The Routine

- For simulated consumer i the market shares are

$$s_{ijt} = \frac{\exp\left(\hat{\delta}^h_{jt} + X_{1jt}(\prod D_i + \sum v_i)\right)}{1 + \sum_k \exp\left(\hat{\delta}^h_{kt} + X_{1kt}(\prod D_i + \sum v_i)\right)}$$

  and we can calculate predicted market shares using

$$\widehat{S_{jt}} = \frac{1}{NS}\sum_i = s_{ijt}$$

- Having a **tight tolerance on the contraction mapping is very important** (e.g., 1e-14).

Random Coefficient Logit Model

Estimation - The Routine

- Often we have a large number of $\prod$, $\sum$ parameters and a search using numerical derivatives will be slow and possibly inaccurate.

- Fortunately analytic derivatives can be calculated and are highly recommended.

- Gradient of the GMM objective function

$$2 * \frac{\partial \xi'(\hat{\theta})}{\partial \theta_{NL}} * W\Phi^{-1}W'\hat{\xi}(\hat{\theta})$$

recall

$$\hat{\xi} = \delta(S, \widehat{\prod}, \widehat{\sum}) - X_1 \underbrace{(X_1'W\Phi^{-1}W'X_1)^{-1}X_1'W\Phi^{-1}W'\delta(S, \widehat{\prod}, \widehat{\sum})}_{\hat{\theta}_1}$$

- Changing the non-linear parameters changes the objective function in two ways: (i) direct effect: changes $\delta(S, \widehat{\prod}, \widehat{\sum})$; (ii) indirect effect: by changing $\delta(S, \widehat{\prod}, \widehat{\sum})$ also changes our estimate of $\hat{\theta}_1$

- It is not too hard to calculate how $\hat{\delta}$ changes with the non-linear parameters using the implicit function theorem

$$F(x_1, x_2, z) = c, \frac{\partial z}{\partial x_1} = -\frac{\frac{\partial F}{\partial x_1}}{\frac{\partial F}{\partial z}}$$

$$S - \hat{S}(\delta, X_1, \widehat{\prod}, \widehat{\sum}) = 0$$

In order to compute the gradient of the objective function, the Jacobian of the function computed in Step 2 has to be computed. This means valuations of the J brands in each market are implicitly defined by the following system of J equations

$$s_j(\delta_1, ..., \delta_j, \theta_2; x, p, P_{ns}) = S_j \qquad j = 1, ..., L.$$

By the Implicit Function Theorem (see Simon and Blume, 1994, Theorem 15.7, page 355) the derivatives of the mean value with respect to the parameters are

$$D\delta = \begin{pmatrix} \frac{\partial \delta_1}{\partial \theta_{21}} & \cdots & \frac{\partial \delta_1}{\partial \theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \delta_J}{\partial \theta_{21}} & \cdots & \frac{\partial \delta_J}{\partial \theta_{2L}} \end{pmatrix} = -\begin{pmatrix} \frac{\partial s_1}{\partial \delta_1} & \cdots & \frac{\partial s_1}{\partial \delta_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_J}{\partial \delta_1} & \cdots & \frac{\partial s_J}{\partial \delta_J} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial s_1}{\partial \theta_{21}} & \cdots & \frac{\partial s_1}{\partial \theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_J}{\partial \theta_{21}} & \cdots & \frac{\partial s_J}{\partial \theta_{2L}} \end{pmatrix}$$

where θ_{2i}, $i = 1, ..., L$ denotes the i's element of the vector θ_2, which contains the non-linear parameters of the model.

$\theta_2 = \theta_{NL}$

Random Coefficient Logit Model

Identification with Aggregate Data and Instruments

- There is a recent literature on non-parametric identification of random coefficient demand models.
  - If you are interested see multiple recent papers by Steve Berry and Phil Haile, and work by Amit Gandhi.
- As a practical matter:
  - Demographic effects identified by how market shares differ across markets with different characteristics.
    - e.g., proportion of kids in UT and FL
    - but often some demographics (e.g., proportion of women) won't vary much
    - or the set of available products will depend on market demographics.

Random Coefficient Logit Model

Identification with Aggregate Data and Instruments

- Random portion identified from how the market shares of different products change in response to price changes or product entry/exit.
    - e.g., if the price of plain Cheerios increases do we observe substitution primarily to products that share a particular characteristic; if so it suggests that there are some consumers who value this characteristic highly
    - or when AC Cheerios is introduced, which products lose share.
- This suggests that we will need a lot of exogenous price variation, exogenous variation in available products and ideally a lot of variation in demographics across markets.

- Recall that we need instruments to identify the non-linear and linear parameters.
  - Even if you treat price as exogenous you still need instruments.
  - We're not just instrumenting for price.
- The choice of instruments can affect the performance of the estimator.

Random Coefficient Logit Model

Instruments

- Various instruments are used in practice:
  - cost-shifters
    - Often hard to find ones that do not shift all products' costs simultaneously.
    - In practice, will want cost-shifters that account for a large proportion of product costs.
  - Hausman style instruments
    - Prices in other markets.
  - characteristics of other products
    - "BLP instruments": e.g., add up the values of an x for other products owned by the same manufacturer and products owned by other manufacturers.
    - These should affect pricing in a Bertrand Nash equilibrium.
    - Requires taking the exogeneity of x very seriously.
- A good idea to interact these variables with demographics?

Random Coefficient Logit Model

Is the BLP/Nevo Algorithm A Good One to Use?

Two obvious criteria:

1. Can it recover the true parameters?
2. Is it quick?

Random Coefficient Logit Model

Can It Recover the True Parameters?

The evidence is that it can recover the true parameters quite reliably as long as:

1. The inner loop (contraction) tolerance is very 'tight' (1e-13).
2. The tolerances (TolX, TolFun) on the numerical minimization (outer loop) are tight (1e-7 or 1e-8).
3. You use analytic derivatives.
4. You use enough simulations to compute shares.
5. You start with fairly good parameter guesses (e.g., you might estimate a logit or nested logit first, or add random coefficients one at a time).
    - A lot of criticisms come from examples where these 5 rules were not used!
        - These examples did lead to people appreciating the importance of these rules.

Random Coefficient Logit Model

Is it Quick/Computationally Efficient?

There's a problem with the algorithm (common to all **nested fixed point** procedures).

- Almost all of the time is spent solving the model for incorrect parameter values!
  - Often for 20 outer loop iterations you are doing millions of inner loop iterations.
    - This is not efficient because you are calculating a lot of exponents.
  - This led to people choosing less accurate inner loop tolerances, and this led to problems ...
- A more efficient algorithm would avoid doing this.
- The most popular approach to trying to do this is to reformulate the problem as a **Mathematical Programming with Equilibrium Constraints** problem.

Random Coefficient Logit Model

MPEC Algorithm (Dube, Fox and Su, Econometrica, 2013)

Under MPEC we solve

$$\min_{\xi,\theta} g(\xi)' W g(\xi) \qquad \text{s.t. } s(\xi;\theta) = S$$

- Now we search over θ and ξ.
  - θ may be e.g., 30×1
  - ξ may be 10,000 or $100,000 \times 1$
- The market share equations only have to hold at the final solution.
- This may look crazy but there are very good methods for solving MPEC problems.
  - Able to exploit sparsity structure in the Jacobian and Hessian (implicit function theorem not required).
  - This is vital for large demand systems with a lot of data.

Random Coefficient Logit Model

MPEC Algorithm (Dube, Fox and Su, forthcoming)

$$\min_{\xi,\theta} g(\xi)' W g(\xi) \qquad \text{s.t. } s(\xi;\theta) = S$$

- This is a **constrained optimization** problem.
- In MATLAB, `fmincon` solves this type of problem (performance questionable).
- For a problem of this type of scale you may need to use routines designed by professionals:
  - SNOPT or KNITRO, which can be run through MATLAB/TOMLAB.
  - AMPL.
  - For examples of the commands required, see Dube or Fox websites.

Random Coefficient Logit Model

How Well Does it Performs?

- MPEC may be useful part of a toolkit.
- But we have not seen a dramatic change in what people do, and it is not obvious how well it works.
- In terms of speed, there is evidence that if it is programmed well it can work significantly faster.
  - It is unclear how this speed up varies with the number of parameters and products.
  - but in most papers you are estimating demand once ...
- However, it is unclear whether the MPEC problem is more or less affected by the problem of having multiple local minima in the objective function.
  - If the problem is worse, this is a significant disadvantage.
  - If the problem is less severe, this is a significant advantage.

Table 1: Three NFP Implementations: Varying Starting Values for One Synthetic Dataset, with Numerical Derivatives

	NFP Loose Inner	NFP Loose Both	NFP Tight	Truth
Fraction Convergence	0.0	0.54	0.95	
Frac. $<1\%>$ "Global" Min.	0.0	0.0	1.00	
Mean Own Price Elasticity	-7.24	-7.49	-5.77	-5.68
Std. Dev. Own Price Elasticity	5.48	5.55	˜0	
Lowest Objective	0.0176	0.0198	0.0169	

We use 100 starting values for one synthetic dataset. The NFP loose inner loop implementation has $\epsilon_{\text{in}} = 10^{-4}$ and $\epsilon_{\text{out}} = 10^{-6}$ The NFP loose-both implementation has $\epsilon_{\text{in}} = 10^{-4}$ and $\epsilon_{\text{out}} = 10^{-2}$. The NFP-tight implementation has $\epsilon_{\text{in}} = 10^{-14}$ and $\epsilon_{\text{in}} = 10^{-6}$. We use numerical derivatives using KNITRO's built-in procedures.

Table 2: Three NFP Implementations: Varying Starting Values for Nevo's Cereal Dataset, with Closed-Form Derivatives

	NFP Loose Inner	NFP Loose Both	NFP Tight Both	NFP Tight Simplex
Fraction Reported Convergence	0.0	0.76	1.00	1.00
Frac. Obj. < 1% Greater than "Global" Min.	0.0	0.0	1.00	0.0
Mean Own Price Elasticity Across All Runs	-3.82	-3.69	-7.43	-3.84
Std. Dev. Own Price Elasticity Across All Runs	0.4	0.07	˜0	0.35
Lowest Objective Function Value	0.00213	0.00683	0.00202	0.00683
Elasticity for Run with Lowest Obj. Value	-6.71	-3.78	-7.43	-3.76

We use the same 50 starting values for each implementation. The NFP loose inner loop implementation has $\epsilon_{\text{in}} = 10^{-4}$ and $\epsilon_{\text{out}} = 10^{-6}$ The NFP loose-both implementation has $\epsilon_{\text{in}} = 10^{-4}$ and $\epsilon_{\text{out}} = 10^{-2}$. The NFP-tight implementation has $\epsilon_{\text{in}} = 10^{-14}$ and $\epsilon_{\text{in}} = 10^{-6}$. The Nelder-Meade or simplex method uses a tighter inner loop tolerance of $\epsilon_{\text{in}} = 10^{-14}$ and MATLAB's default values for the simplex convergence criteria. We manually code closed-form derivatives for all methods other than for Nelder-Meade, which does not use derivative information.

Table 5: Monte Carlo Results Varying the Number of Markets, Products and Simulation Draws

# Markets T	# Products J	# Draws n_s	Lipsch. Const.	Imple.	Runs Converged	CPU Time (hours)	Outside Share
100	25	1000	0.999	NFP tight	80%	10.9	0.45
				MPEC	100%	0.3	
250	25	1000	0.997	NFP tight	100%	22.3	0.71
				MPEC	100%	1.2	
500	25	1000	0.998	NFP tight	80%	65.6	0.65
				MPEC	100%	2.5	
100	25	3000	0.999	NFP tight	80%	42.3	0.46
				MPEC	100%	1	
250	25	3000	0.997	NFP tight	100%	80	0.71
				MPEC	100%	3	
25	100	1000	0.993	NFP tight	100%	5.7	0.28
				MPEC	100%	0.5	
25	250	1000	0.999	NFP tight	100%	28.4	0.07
				MPEC	100%	2.3	

There is one dataset and five starting values for each experiment. The mean intercept is 2. MPEC and NFP produces the same lowest objective value.

Berry, Levinsohn and Pakes (1995)

Setting and Data

- Estimate demand for automobiles.
  - This was very clearly a methodology paper.
  - They had some related papers looking at economic questions such as the impact of import restraints from Japan.
- A product is a model, and a market is one year of data in the US, so we do not have much.
- Data:
  - price: list retail price for base model in 1983 $s
  - quantity: US sales by name plate of new cars
  - market size: number of households
  - characteristics: cylinders, weight, MP$, HPW etc
  - observe about 150 models per year

Table I
Descriptive Statistics

Year	No. of Models	Quantity	Price	Domestic	Japan	European	HP/Wt	Size	Air	MPG	MP$
1971	92	86.892	7.868	0.866	0.057	0.077	0.490	1.496	0.000	1.662	1.850
1972	89	91.763	7.979	0.892	0.042	0.066	0.391	1.510	0.014	1.619	1.875
1973	86	92.785	7.535	0.932	0.040	0.028	0.364	1.529	0.022	1.589	1.819
1974	72	105.119	7.506	0.887	0.050	0.064	0.347	1.510	0.026	1.568	1.453
1975	93	84.775	7.821	0.853	0.083	0.064	0.337	1.479	0.054	1.584	1.503
1976	99	93.382	7.787	0.876	0.081	0.043	0.338	1.508	0.059	1.759	1.696
1977	95	97.727	7.651	0.837	0.112	0.051	0.340	1.467	0.032	1.947	1.835
1978	95	99.444	7.645	0.855	0.107	0.039	0.346	1.405	0.034	1.982	1.929
1979	102	82.742	7.599	0.803	0.158	0.038	0.348	1.343	0.047	2.061	1.657
1980	103	71.567	7.718	0.773	0.191	0.036	0.350	1.296	0.078	2.215	1.466
1981	116	62.030	8.349	0.741	0.213	0.046	0.349	1.286	0.094	2.363	1.559
1982	110	61.893	8.831	0.714	0.235	0.051	0.347	1.277	0.134	2.440	1.817
1983	115	67.878	8.821	0.734	0.215	0.051	0.351	1.276	0.126	2.601	2.087
1984	113	85.933	8.870	0.783	0.179	0.038	0.361	1.293	0.129	2.469	2.117
1985	136	78.143	8.938	0.761	0.191	0.048	0.372	1.265	0.140	2.261	2.024
1986	130	83.756	9.382	0.733	0.216	0.050	0.379	1.249	0.176	2.416	2.856
1987	143	67.667	9.965	0.702	0.245	0.052	0.395	1.246	0.229	2.327	2.789
1988	150	67.078	10.069	0.717	0.237	0.045	0.396	1.251	0.237	2.334	2.919
1989	147	62.914	10.321	0.690	0.261	0.049	0.406	1.259	0.289	2.310	2.806
1990	131	66.377	10.337	0.682	0.276	0.043	0.419	1.270	0.308	2.270	2.852
All	2217	78.804	8.604	0.790	0.161	0.049	0.372	1.357	0.116	2.099	2.086

Note: The entry in each cell of the last nine columns is the sales weighted mean.

- Utility:
$$U_{ij} = X_j \beta_i + \alpha \ln(y_i - p_j) + \xi_j + \varepsilon_{ij}$$
the α parameter is common across people, but variation in income y_i will provide variation in price sensitivity.

- BLP use the distribution of income in the US to give y_i.

- This is an appealing approach for expensive products.

- No demographics apart from income (which only affects the price coefficient).

Berry, Levinsohn and Pakes (1995)

Instruments

- There is a long discussion of what makes an optimal instrument.
- in practice they are going to use

$$z_{jk}, \sum_{r \neq j, r \in \mathscr{F}_f} z_{rk}, \sum_{r \neq j, r \notin \mathscr{F}_f} z_{rk}.$$

  for different sets of characteristics (includes a constant, so e.g. the count of the number of products owned).
- Faced some collinearity issues (e.g., MP$ and MPG) so some are dropped.
- In total 34 moments and 17 or 18 parameters.

Berry, Levinsohn and Pakes (1995)
Supply

- BLP also **impose the supply-side** during estimation.
  - Both demand and supply are static.
  - i.e., optimal price setting (not optimal product selection).
- marginal cost

$$mc_j = \exp(X_j \gamma + \omega_j)$$

- the problem of a multi-product firm:

$$\max \pi_{ft} = \sum_{j \in F_f} (p_{jt} - mc_{jt}) M s_{jt}(p)$$

Supply Side

First Order Conditions

Assuming that all prices are positive

$$s_{jt}(p) + \sum_{r \in F_f} (p_{rt} - mc_{rt}) \frac{\partial s_{rt}(p)}{\partial p_{jt}} = 0$$

$$S_{jr} = -\frac{\partial s_r}{\partial p_j}, H = \text{ownership matrix}$$

$$H_{jr} = 1 \text{ iff products } j \text{ and } r \text{ have same owner}$$

$$\Omega_{jr} = H_{jr} * S_{jr}$$

$$p - mc = \Omega^{-1} s(p)$$

- BLP's specification is that $\ln(mc) = w\gamma + \omega$

$$\ln \underbrace{\left(p - \Omega^{-1} s(p)\right)}_{\text{must be possitive}} = w\gamma + \omega.$$

- The non-linear demand parameters will show up in $\frac{\partial s_{rt}(p)}{\partial p_{jt}}$.

Supply Side

Incorporation into Estimation

- Form additional moments using the unobserved component of marginal costs (here the structural error)

$$\ln\left(p - \Omega^{-1} s(p)\right) - w\gamma = \omega.$$

- Own characteristics are treated as exogenous, and, once again, use characteristics of other products as instruments.

BLP

Objective Function

- The model is estimated using GMM, where we stack the demand and supply equations.
- They factor out the linear parameters, so that they only need to search over the non-linear parameters.
    - They have 6 of these (all cost parameters are linear).
- They don't use derivatives; instead use Nelder-Mead.
- In general the derivatives will become more complicated when the additional supply side moments are imposed.

Berry, Levinsohn and Pakes (1995)

Estimates: Comparing OLS Logit, IV Logit

- OLS implies that most models have inelastic demand.
  - Inconsistent with profit maximization (esp. as multi-product concerns should lead to even higher prices).
- IV Logit (estimated using the Berry inversion)
  - Almost all products have elastic demands.
  - But all models have small market shares and almost identical mark-ups ($\$4, 700 \pm 100$).
  - Firms with larger product portfolios have larger mark-ups.
    - Seems a problem as GM, Ford have bigger portfolios than Ferrari and BMW.

Results with Logit Demand and Marginal Cost Pricing
(2217 Observations)

Variable	OLS Logit Demand	IV Logit Demand	OLS ln (price) on *w*
Constant	-10.068	-9.273	1.882
	(0.253)	(0.493)	(0.119)
HP/ Weight*	-0.121	1.965	0.520
	(0.277)	(0.909)	(0.035)
Air	-0.035	1.289	0.680
	(0.073)	(0.248)	(0.019)
MP$	0.263	0.052	-
	(0.043)	(0.086)	
MPG*	-	-	-0.471
			(0.049)
Size*	2.341	2.355	0.125
	(0.125)	(0.247)	(0.063)
Trend	-	-	0.013
			(0.002)
Price	-0.089	-0.216	-
	(0.004)	(0.123)	
No. Inelastic Demands	1494	22	n.a.
(+ / - 2 s.e.'s)	(1429-1617)	(7-101)	(aaa)
R^2	0.387	n.a.	0.656

Berry, Levinsohn and Pakes (1995)

Estimates: Random Coefficient Model

- All models have elastic demand.
- More crowded parts of the product space (e.g., compacts) have higher elasticities.
- Substitution to the outside goods varies across cars (it's higher for cheaper cars).
- Mark-ups much more reasonable, and vary a lot across products.

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

TABLE V

A SAMPLE FROM 1990 OF ESTIMATED DEMAND ELASTICITIES WITH RESPECT TO ATTRIBUTES AND PRICE (BASED ON TABLE IV (CRTS) ESTIMATES)

	Value of Attribute/Price Elasticity of demand with respect to:				
Model	*HP/Weight*	*Air*	*MP\$*	*Size*	*Price*
Mazda323	0.366	0.000	3.645	1.075	5.049
	0.458	0.000	1.010	1.338	6.358
Sentra	0.391	0.000	3.645	1.092	5.661
	0.440	0.000	0.905	1.194	6.528
Escort	0.401	0.000	4.022	1.116	5.663
	0.449	0.000	1.132	1.176	6.031
Cavalier	0.385	0.000	3.142	1.179	5.797
	0.423	0.000	0.524	1.360	6.433
Accord	0.457	0.000	3.016	1.255	9.292
	0.282	0.000	0.126	0.873	4.798
Taurus	0.304	0.000	2.262	1.334	9.671
	0.180	0.000	−0.139	1.304	4.220
Century	0.387	1.000	2.890	1.312	10.138
	0.326	0.701	0.077	1.123	6.755
Maxima	0.518	1.000	2.513	1.300	13.695
	0.322	0.396	−0.136	0.932	4.845
Legend	0.510	1.000	2.388	1.292	18.944
	0.167	0.237	−0.070	0.596	4.134
TownCar	0.373	1.000	2.136	1.720	21.412
	0.089	0.211	−0.122	0.883	4.320
Seville	0.517	1.000	2.011	1.374	24.353
	0.092	0.116	−0.053	0.416	3.973
LS400	0.665	1.000	2.262	1.410	27.544
	0.073	0.037	−0.007	0.149	3.085
BMW 735i	0.542	1.000	1.885	1.403	37.490
	0.061	0.011	−0.016	0.174	3.515

Notes: The value of the attribute or, in the case of the last column, price, is the top number and the number below it is the elasticity of demand with respect to the attribute (or, in the last column, price.)

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.

TABLE VII

SUBSTITUTION TO THE OUTSIDE GOOD

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

TABLE VIII

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

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