Empirical Models of Product Differentiation

Steven Berry

Yale Univ., Dept. of Economics and Cowles Foundation
and NBER

November 14, 2007

Introduction

Many markets are well-described by models that feature:

- Oligopolistic firms
- Differentiated Products
- Competition in
  - price
  - quality, characteristics, location, etc.

Idea of Talks

How do we estimate the parameters of such models from real world data, and how do we use the estimates to analyze interesting policy questions?

- Talk 1: Differentiated Products Demand Estimation
- Talk 2: Product Choice and Variety
- Talk 3: Policy Applications

Today's Outline

Intro to Diff Products

While homogeneous goods is a convenient assumption for many models, it is frequently violated in practice.

Many markets feature goods that vary in

- quality,
- physical characteristics,
- reliability,
- reputation,
- geographic location.

"Policy" Questions

- Mergers
- Marketing, pricing, etc.
- Tax effect / incidence
- Welfare from new products
- "True" CPI
- "Nature" of competition

Example: Anti-Trust

Homogeneous Goods

- Market Definition
- "In or out" of market
- Level of concentration
- Ease of entry

Example: Anti-Trust

Homogeneous Goods

- Market Definition
- "In or out" of market
- Level of concentration
- Ease of entry

Diff. Products

- Degree of substitutability
- Pattern of joint ownership
- Entry into space of characteristics

Theories of Product Differentiation

Consumers

Preferences for

- Products, or
- Characteristics of products
  - "address" models (Hotelling),
  - "Characteristics" models (Vertical, Lancasterian)

Theories of Product Differentiation

Consumers

Preferences for

- Products, or
- Characteristics of products
  - "address" models (Hotelling),
  - "Characteristics" models (Vertical, Lancasterian)

We can have either

- Discrete Choice, or
- Continuous Choice, or
- Mixed choice

Demand-Side Reasons for Product Diff

Variety

Declining MU for each good, often representative consumer, continuous choice (e.g. Dixit-Stiglitz)

Demand-Side Reasons for Product Diff

Variety

Declining MU for each good, often representative consumer, continuous choice (e.g. Dixit-Stiglitz)

Heterogeneous Tastes

Often pure discrete choice, both classic theory models (Hotelling) and classic empirical models (McFadden)

Demand-Side Reasons for Product Diff

Variety

Declining MU for each good, often representative consumer, continuous choice (e.g. Dixit-Stiglitz)

Heterogeneous Tastes

Often pure discrete choice, both classic theory models (Hotelling) and classic empirical models (McFadden)

How to combine?

Ideal model might have both a "taste for variety" and also heterogeneous tastes.

Empirical Approaches

- Direct Tastes for Products (AIDS) vs. Discrete Choice
- Market Level Data vs. Consumer Level Data vs. Panel on Consumers across different Markets
- Or, combinations of these

Direct Tastes for Products

Good for empirical work if you have

- The same small set of products
- offered in many market

Can model with convenient function form (log-linear, Almost Ideal, etc.)

$$q_{jt} = f(p_j, p_{-j}, \theta) + \epsilon_j$$

Need instruments for price. See, e.g. Hausman (1996) [8].

Direct Tastes for Products

Good for empirical work if you have

- The same small set of products
- offered in many market

Can model with convenient function form (log-linear, Almost Ideal, etc.)

$$q_{jt} = f(p_j, p_{-j}, \theta) + \epsilon_j$$

Need instruments for price. See, e.g. Hausman (1996) [8].

But,

if we have many products per market, then have "too many elasticities" to estimate. And how to deal with product choice sets that change across markets?

Discrete Choice Models

- Differentiated by an explicit set of characteristics
- Utility depends on characteristics of good interacted with attributes of consumer
- Purchase decision is one or none
- Data include qty, price and characteristics,
- Cross-section of markets,
- Some information on consumer choices,
- Prices are set by multi-product oligopolists,
- Characteristics not endogenous until tomorrow!

Utility Specification

Utility

$$U(x_{jt}, \xi_{jt}, z_{it}, \nu_{it}; \theta_d),$$

where

- x_{jt} is observed product characteristic,
- ξ_{jt} is unobserved (by us) product characteristic,
- z_{it} is observed consumer attribute (income, age, ...)
- ν_{it} is unobserved consumer attribute ("random taste")
- θ_d is demand parameter

Utility Specification

Utility

$$U(x_{jt}, \xi_{jt}, z_{it}, \nu_{it}; \theta_d),$$

where

- x_{jt} is observed product characteristic,
- ξ_{jt} is unobserved (by us) product characteristic,

Product characteristics

Utility Specification

Utility

$$U(x_{jt}, \xi_{jt}, z_{it}, \nu_{it}; \theta_d),$$

where

- z_{it} is observed consumer attribute (income, age, ...)
- ν_{it} is unobserved consumer attribute ("random taste")

Consumer attributes

Utility Specification

Utility

$$U(x_{jt}, \xi_{jt}, z_{it}, \nu_{it}; \theta_d),$$

where

- x_{jt} is observed product characteristic,
- z_{it} is observed consumer attribute (income, age, ...)

Observables

Utility Specification

Utility

$$U(x_{jt}, \xi_{jt}, z_{it}, \nu_{it}; \theta_d),$$

where

- ξ_{jt} is unobserved (by us) product characteristic,
- ν_{it} is unobserved consumer attribute ("random taste")

Unobservables

Utility Specification

Utility

$$U(x_{jt}, \xi_{jt}, z_{it}, \nu_{it}; \theta_d),$$

where

- θ_d is demand parameter

Parameters

Unobserved Product Characteristic, ξ

.

Important because ...

- Realistic that x's are missing
- Explains why shares don't fit exactly ("residual")
- Implies endogeneity of price (and x?)

Choice Data

Market Level Data

We see only product-level market shares

$$s_{jt} = q_{jt}/M_t$$

for a cross-section of markets

Choice Data

Market Level Data

We see only product-level market shares

$$s_{jt} = q_{jt}/M_t$$

for a cross-section of markets

Consumer Level Data

We also see the match between consumers and their choices: y_{it} is an index of consumer i's choice. Or, we might see less detailed information (e.g. the average income of purchasers of good j.) In unusual cases, see second-choice data and/or multiple choices by the same consumer.

Endogeneity

Early discrete choice models ignored endogeneity because estimated on consumer-level data. We will see that this is wrong.

Either there is an product- or market-level unobservable (ξ) that is correlated with price (or other x's) or else there is not. This does not depend on the aggregation of the data.

When there is an unobservable, we face the problem of how to construct a *ceteris paribus* counterfactual – we want to know what happens when we change price, holding ξ constant. This is a special problem when p and ξ are correlated.

Consumer Level Data

See McFadden (1977) [10] as modified by Berry (1994) [1], "BLP" (1994) [2], and "MicroBLP (2004)" [3].
With consumer ("micro") data, consumer purchase probability for product j is

$$Prob(y_{it} = j \mid z_{it}, x_t, \xi_t \mid \theta_d) \equiv P_j(z_{it}, x_t, \xi_t \mid \theta_d)$$

$$= \int_{A_j(z_{it}, x_t, \xi_t \;\; ; \theta_d)} \Phi(d\nu)$$

where:

$$A_j(z, x, \xi; \theta_d) =$$

$$[\nu : U(x_j, \xi_j, z, \nu; \theta_d) \geq U(x_k, \xi_k, z, \nu; \theta_d), \quad \forall k \neq j]$$

Assumptions on Consumer Random Taste, ν_i

- Pure Logit
- Nested Logit (McFadden (1978) [9], Cardell (1977) [6])
- Random Coefficients Logit (McFadden, Hausman and Wise (1978) [7])
- "Pure Characteristics" Berry and Pakes (2007) [4],
- Non-parametric (Berry-Haile, in progress)

Logit

Random consumer tastes are:

$$\nu_{it} \equiv (\epsilon_{i0t}, \epsilon_{i1t}, \ldots, \epsilon_{iJt}),$$

where the ϵ's are i.i.d. "double exponential", and the utility is

$$u_{ijt} = \delta_{jt} + \sum_k \sum_r z_{ikt} x_{jrt} \gamma_{rk} + \epsilon_{ijt}$$

$$= \delta_{jt} + z_{it}' \Gamma x_{jt} + \epsilon_{ijt}$$

Logit

Random consumer tastes are:

$$\nu_{it} \equiv (\epsilon_{i0t}, \epsilon_{i1t}, \ldots, \epsilon_{iJt}),$$

where the ϵ's are i.i.d. "double exponential", and the utility is

$$u_{ijt} = \delta_{jt} + \sum_k \sum_r z_{ikt} x_{jrt} \gamma_{rk} + \epsilon_{ijt}$$

$$= \delta_{jt} + z_{it}' \Gamma x_{jt} + \epsilon_{ijt}$$

The "product-specific constant" (Berry, 1994) is

$$\delta_{jt} \equiv x_{jt}\beta + \xi_j$$

Estimation of the Logit

on Micro Data

- Very easy to estimate (δ,γ) from MLE:

$$\frac{e^{\delta_{jt}+z_{it}'\Gamma x_{jt}}}{\sum_r e^{\delta_{rt}+z_{it}'\Gamma x_{rt}}}$$

Estimation of the Logit

on Micro Data

- Very easy to estimate (δ,γ) from MLE:

$$\frac{e^{\delta_{jt}+z_{it}'\Gamma x_{jt}}}{\sum_r e^{\delta_{rt}+z_{it}'\Gamma x_{rt}}}$$

- Now, are left with the "aggregate" problem:

$$\delta_{jt} = x_{jt}\beta + \xi_j$$

Consumer data identifies γ and δ, but doesn't say how x and ξ shift δ. Also, doesn't "solve" the endogeneity problem.

Price Endogeneity

Introduce price as a separate characteristic:

$$\delta_{jt} = \bar{x}_{jt}\bar{\beta} - \alpha p_{jt} + \xi_{jt}$$

We need cross-product variation in x and instruments for price. (Probably, ξ is correlated with p.)
Price instruments could be:

- Cost shifters
- Mark-up shifters

One source of variation in mark-ups: cross-market changes in choice sets (effects markups.)

"Panel Data" Identification

Perhaps ξ_{jt} is constant for some product group (across markets, or across "sub-products" of a firm.) In this case, can get price coefficient from "within product group" variation in price.

With many consumers per product, can probably reject the assumption (because won't correctly predict total product shares.)

Restrictiveness of Logit

With no variation in consumer tastes (other than ϵ), the logit own- and cross-product demand derivatives are

$$\frac{\partial s_j}{\partial p_k} = \alpha s_k s_j.$$

Substitution is to popular, not similar, products.
If choices vary with z_i, then

$$\frac{\partial s_j}{\partial p_k} = \int \alpha(z) s_j(z) s_k(z) dF(z)$$

If z's shift market shares "enough" can cause substitution to similar products – but may be insufficient.

Random Coefficients Logit

Consider

$$u_{ijt} = x_j \beta_i + \epsilon_{ijt}$$

$$\beta_{ir} = \sum_k z_{ik} \gamma_{rk} + \sigma_r \nu_{ir},$$

$$\nu_{ir} \sim N(0, \sigma_r^2)$$

Larger variance (σ_r^2) in the "Random coefficients" leads to tighter substitution patterns.

"Identification" of σ

In general, identification of demand comes from confronting the "same preferences" (demand parms) with "different choices."

Information on the nature of substitution patterns can come from

- changes in the choice set across markets,
- and/or information on second choices and/or
- repeated choices

Instruments capturing "changes in choice set"

- Functions of other product product's characteristics
- With many products per market, interact own-x with a market (x has "different effect" in markets with different
- "Optimal Instruments" predict $\partial \xi / \partial \theta$.

Estimation Issues with Random Coefficients

- No closed form for shares: have to simulate
- Simulation of small shares can be a problem for MLE. GMM may be better, but still an issue.
- As in MicroBLP (2004) can fit δ to total market shares, if avail.
- OK if consumer sample is choice-based or missing some products (use model to correct likelihood.)

Semi-Parametric Identification

Berry & Haile, in progress

One possible criticism of "structural model": are we imposing an answer by choosing the parametric assumption. With logit, maybe yes!

At at least in principle, would like to know if it is possible to uncover the distribution of random utility, conditional on x, ξ and z, with no parametric assumptions on ν and β.

Novel part here is the unobservable ξ.

Assume for each i there is some monotonic transformation of utility such that

$$\tilde{u}_{ijt} = \beta_i z^1_{ijt} + \mu\left(x_{jt}, \xi_{jt}, z^2_{ijt}, \nu_i\right) \qquad \forall i, j = 1, \ldots, \mathcal{J}^t$$

Assume for each i there is some monotonic transformation of utility such that

$$\tilde{u}_{ijt} = \beta_i z^1_{ijt} + \mu\left(x_{jt}, \xi_{jt}, z^2_{ijt}, \nu_i\right) \qquad \forall i, j = 1, \ldots, \mathcal{J}^t$$

which is further transformed to

$$u_{ijt} = z^1_{ijt} + \frac{\mu\left(x_{jt}, \xi_{jt}, z^2_{ijt}, \nu_i\right)}{\beta_i} \qquad \forall i, j = 1, \ldots, \mathcal{J}^t.$$

$$\equiv z^1_{it} + \delta_i\left(x_t, \xi_t\right),$$

$$\equiv z^1_{it} + \delta^t_i.$$

Binary Choice (1 Product) Example

Choose the good when:

$$z^1_{it} + \delta^t_i > 0.$$

First, consider identification under the following conditions.

I.i.d. random tastes

$(\nu_i, \beta_i) \perp (x_t, z_{ijt}, \xi_t)$.

Full support on z^1

supp $z^1_{it} | t = \mathbb{R} \; \forall t$.

Step 1: Identification of the Within Market Taste Dist.

It is well know that

$$\Pr\left(y_{it} = 1 \mid x_t, w_{it}\right) = \Pr\left(\delta_{it} \leq -z_{it}\right)$$

the distribution of δ_{it} is identified in each market from the variation in choices across z^1.

But we want to know how this distribution varies with x and ξ – and we don't know ξ.

Inferring "mean tastes"

Idea (similar to logit), infer ξ from mean tastes,

$$\bar{\delta}^t = E\left[\delta_i^t(x_t, \xi_t, \nu_i)\right]$$

$$= g\left(x_t, \xi_t\right)$$

for some g that is strictly increasing in its second argument.

From prior step, we know $\bar{\delta}^t$ as mean of δ_{it}.

Or, might uncover median instead (need additive ξ now.)

Step 2: Uncovering ξ

For each (j, t) we have the "mean taste" equation,

$$\bar{\delta}^t = g\left(x_t, \xi_t\right)$$

which is easily identified if x and ξ are independent. If some x (price?) is endogenous, then one could from the beginning assume that ξ enters utility additively and so the mean taste is

$$\bar{\delta}^t = g(x_t) + \xi_t$$

and "semi-parametric IV" (Newey-Powell) gives g and ξ. Also, if z's have a dense, but not full support, can still use quantile methods to uncover ξ. (Have to learn some quantile in every market.)

Step 3: Conditional Dist. of Tastes

Now that we know ξ and the tastes in each market, in principle can uncover, from enough cross-market data, how the full distribution of tastes varies with x and ξ.

This is a generalization of the random coefficient model.

Extension to Many Goods

- Now need different z_j^1, one for each product, that move separately to identified within market tastes. Example: distance from consumer to product. "Full support" is even less reasonable now.
- But given full support can then get mean utility levels and proceed as before.
- Extension can handle dense support over some region in R^J, but need overlap across markets in the region of identified tastes.

Practical Application

In practice, will probably still use some parametric assumptions in the case with many goods, because of

1. problems with support of z
2. usual problem of dimensionality.

Can choose where parametric assumption is most useful, depending on problem. Also, get "goodness of fit" to non-parametric model.

Estimation on Market-Level Data

As in BLP [2]:

$$u_{ijt} = \delta_j + \left[\sum_r x_{jr} \sigma_r \nu_{ir} \right] + \epsilon_{ijt}$$

$$\nu_{ir} \sim N(0, \sigma_r^2)$$

$$\delta_j = x_j \beta + \xi_j$$

All the same issues as with micro data, but now don't have the z_i's to "sweep out" the correct distribution of random tastes. So, have to make a parametric assumption on the ϵ's. Still: key idea is for the function form to be "flexible enough" to match interesting patterns of substitution (e.g. *not* the pure logit.)

Identifying σ

Again, from cross-market variation in choice sets. This is similar to the "second stage" of the problem with consumer data.

Note: discrete choice model chooses a functional form, but in general we think market-level demand may be identified from sufficient exogenous variation in choice sets.

Idea of Estimation

$$s_j(\delta, x, \sigma) = q_j/M$$

Can "invert" this to uncover

$$\delta_j(s, q_j/M, \sigma) = x_j\beta + \xi_j$$

Want instruments that satisfy

$$E(\xi_{jt} \mid z_{jt}) = 0$$

or else "panel data" restrictions on ξ.

Example: Estimation of the Logit

On Market Data

Data

Market qty, price, characteristics. Market share $s_j = q_j/M$.

Market Share

$$s_j = \frac{e^{\delta_j}}{1 + \sum_{r=1}^{J} e^{\delta_r}},$$

get

$$\begin{aligned} ln(s_j) - ln(s_0) &= \delta_j - \delta_0 \qquad (1) \\ &\equiv x_j\beta - \alpha p_j + \xi_j. \end{aligned}$$

Can now estimate by 2SLS

Extensions to Logit

Nested logit McFadden (1978) ([9], Cardell (1997) [6]): logit within groups (nests) and "logit-like" choice of groups, but within group choices are more correlated than across group choices.
For product j in group g:

$$u_{ij} = x_j\beta - \alpha p_j + \xi_j + \sigma_g \nu_{ig} + \epsilon_{ij}.$$

New "within-group correlation" parameter is σ_g.

Related to this: GEV models (e.g. Bresnahan, Stern and Trachtenberg (1997), [5])

Also: mixtures of logits (K "types" of logit parameters).

Random Coefficients on Market Level Data

Extending the Logit Method

With random coefficients, can still solve for δ_j from market share equations, and estimate parameters of δ by IV.
But now also need "instruments" (i.e. moment restrictions) for the substitution patterns.

Berry, Levinsohn and Pakes '95 (BLP) provide a simple algorithm for solving the share equations for δ. Many computational details ...

The GMM (generalized method of moments) estimation algorithm:

- Guess a parameter
- Solve for δ and therefore ξ.
- interact ξ and instruments z – these are the moment conditions $G(\theta)$.
- Calculate an objective function – how far is $G(\theta)$ from zero? $f(\theta) = G'AG$ for some positive definite A.
- Guess a new parameter and try to minimize f.
- Variance of $\hat{\theta}$ includes variance in data across products and simulation error as well as any sampling variance in the observed market shares.

(Can simplify the algorithm since δ in linear in some parameters.)

"Identification"

Need changes in choice set over time to identify substitution pattern.
In practice:

- cost shifters (price changes),
- "characteristics of other goods",
- interactions between x and market dummy,
- Panel Data: restriction on how ξ_{jt} changes over time.

Question:

Is there enough variance to identify substitution from demand alone? Solutions: more markets, consumer ("micro") data and/or add a "supply" side.

Pure Random Coefficients

Berry and Pakes (2007) consider the "no epsilon" model:

$$u_{ijt} = \delta_j + \left[\sum_r x_{jr} \sigma_r \nu_{ir} \right] + \epsilon_{ijt}$$

This is more similar to the theory literature (Hotelling, Salop, etc.) where there is typically not an "i.i.d." match value.

Pure Random Coefficients

Berry and Pakes (2007) consider the "no epsilon" model:

$$u_{ijt} = \delta_j + \left[\sum_r x_{jr} \sigma_r \nu_{ir} \right] + \epsilon_{ijt}$$

This is more similar to the theory literature (Hotelling, Salop, etc.) where there is typically not an "i.i.d." match value.

- Can still invert for δ
- Inversion is computationally harder
- Problem of simulation is s not as bad
- Special case of the model with ϵ, but at the limit with $\mathrm{Var}(\epsilon) = 0$.

Non-parametric version

BLP and Berry-Pakes '07 [4] prove the existence of an "inverse" share equation that uncovers the unobserved product characteristeics":

$$\xi_j = f(s_j, s_{-j}, x_j, x_{-j}, \theta)$$

This looks like literature on estimation of non-parametric models with "non-separable" errors (e.g. Chernozukov and Hansen.)

Open question about conditions for non-parametric identification of this model.

Partial Consumer Data

Often, have some limited data on consumers: tabulation of some consumer means, data on some products only, etc.

Advantage of model is can always make use of this data.

Example: additional moment is that model must predict average income of purchasing consumers in each market.

Example: Petrin

Petrin's '02 [11] asked about the welfare of the introduction of the minivan.

From market-level data, poor estimates of the substitution within minivans. Add: demographics of minivan buyers. "Two or more children" is strong predictor and adding this greatly increasing the estimated substitutability across minivans.

Conclusions

- We want to estimate a rich set of differentiated product models in order to do applications and policy analysis.
- On consumer level data, can estimate
  - Effect of consumers attributes from consumer choice data
  - Need choice-set variation to get further info on substitution patterns,
  - Need exogenous variation in costs and/or choice sets to solve endogeneity problem,
  - Variation in consumer-level data can provide non-parametric identification of random tastes within market; and then cross-market variation identifies effect of unobservables and, then, how the distribution of tastes changes with x and ξ.
- With market-level data, need some parametric assumptions, but still note the important of cross-market variation in choices.
- Extensions include multiple choices and the "pure characteristics" model.

Next

- Endogenous Product Characteristics
  - Static
  - Dynamic
- Applications and Policy

Steven Berry.
Estimating discrete choice models of product differentiation.
RAND Journal of Economics, 23(2):242–262, Summer 1994.

Steven Berry, James Levinsohn, and Ariel Pakes.
Automobile prices in market equilibrium.
Econometrica, 60(4):889–917, July 1995.

Steven Berry, James Levinsohn, and Ariel Pakes.
Differentiated products demand systems from a combination of micro and macro data: The new vehicle market.
Journal of Political Economy, 112(1):68–105, February 2004.

Steven T. Berry and Ariel Pakes.
Estimating the pure characteristics model of discrete-choice demand.

International Economics Review (Special Issue in Honor of Dan McFadden, 2007.

Timothy Bresnahan, Scott Stern, and Manuel Trajtenberg.
Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s.
RAND Journal of Economics, Special Issue, 28(0):S17–44, 1997.

N. Scott Cardell.
Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity.
Econometric-Theory, 13(2):185–213, April 1997.

J.A. Hausman and D. Wise.
A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences.
Econometrica, 46:403–426, 1978.

Jerry A. Hausman.
Valuation of new goods under perfect and imperfect competitioin.
In Timothy F. Bresnahan and Robert J. Gordon, editors, *The Economics of New Goods*, chapter 5, pages 209–248. University of Chicago Press, Chicago, 1996.

Daniel McFadden.
Modelling the choice of residential location.
In A. Karlvist, editor, *Spatial Interaction Theory and Planning Models*, pages 75–96. North Holland, Amsterdam, 1978.

Daniel McFadden, A. Talvitie, and Associates.
Demand Model Estimation and Validation.
Institute of Transportation Studies, Berkeley CA, 1977.

Amil Petrin.
Quantifying the benefits of new products: The case of the minivan.
JPE, 110(4):705–729, August 2002.