Discrete Choice Models and the Demand for Differentiated Products: BLP and BFM

Holger Sieg
University of Pennsylvania

How Do We Estimate Demand Models?

- The study of demand is perhaps the most common example of structural modeling in empirical microeconomics.
- The basic goal: estimate the own- and cross-price elasticities for the set of goods in the market under study.
- One common problem in most settings is that there are too many goods.
- In many product markets, there are many varieties of products.
- This means there are many many own- and cross-price elasticities.
- In practice, there are often too many goods to practicably estimate price elasticities without imposing additional restrictions.

Aggregate Demand

- To see why this is a problem, consider estimating a simple linear demand system:

$$
\begin{aligned}
q_1 &= \beta_{0,1} + \beta_{1,1} p_1 + \cdots + \beta_{J,1} p_J + \varepsilon_1 \\
&\vdots \\
q_J &= \beta_{0,J} + \beta_{1,J} p_1 + \cdots + \beta_{J,J} p_J + \varepsilon_J
\end{aligned}
$$

- With J products, there are J^2 elasticities to estimate.
- Thus with J=140 types of butter and margarine, there are almost 20,000 elasticities to estimate.
- Even *without* endogeneity concerns, one would need *lots* of data to try to pin down so many parameters.
- We need to impose some structure on the problem.

The Problem of Differentiated Products

- Almost all consumer goods are not homogeneous goods.
- Instead they are differentiated products that differ along observed and potentially unobserved dimensions.
- Product differentiation provides limited market power that allows firms to set prices exceeding marginal costs. This is necessary to recover fixed costs.
- We use a characteristics approach that goes back to Gorman and Lancaster.
- We treat a product as a bundle of characteristics that can be described by a finite dimensional vector of attributes.
- We then aggregate the individual demand functions to generate the market level demand.

Discrete Choice Fundamentals

Consider a general specification of a discrete-choice problem:

- A consumer:

$$\max_{q_1, q_2, c} U(q_1, q_2, c)$$

  subject to

$$\begin{aligned} p_1 q_1 + p_2 q_2 + c &= m \\ q_1 \; q_2 &= 0 \end{aligned}$$

  where
  - q_1 and q_2 are the quantities of the discrete alternatives the consumer is choosing,
  - c is the numeraire good (with its price normalized to 1),
  - m is income.
- The second constraint embodies the discreteness of the decision: either the consumer buys good 1 or she buys good 2, but not both.

Conditional Indirect Utility Functions

- Suppose that we *condition* on $q_1 = 0$.
- In this case, it is a standard utility maximization problem:

$$\max_{q_2, c} U(0, q_2, c)$$

subject to

$$p_2 \; q_2 + c = m$$

- There is a standard solution for the demand functions $q_2(p_2, m)$ and $c(p_2, m)$. Plugging these demand functions back into the utility function yields the the conditional indirect utility function:

$$V_2(p_2, m) = U(0, q_2(p_2, m), c(p_2, m))$$

Optimal Choices

- Doing the same *conditioning* on $q_2 = 0$, we obtain an analogous expression:

$$V_1(p_1, m) = U(q_1(p_1, m), 0, c(p_1, m))$$

- The solution to the original decision problem is then given by a discrete choice between the conditional indirect utility functions:

$$\max_{j=\{1,2\}} V_j(p_j, m)$$

- Dubin and McFadden (1984) discuss how to estimate these mixed discrete-continuous choice models.You need two errors one for the discrete part of the choice problem and one for the continuous choice.

Pure Discrete Choice Models

- The problem becomes significantly easier if we assume that $q_1, q_2 \in \{0, 1\}$, i.e quantities can only be zero or one.
- For example, you can only live in one city, buy one car, live in one house, or go on one vacation at a time.
- In that case, we have:

$$\begin{aligned} V_1(p_1, m) &= U(1, 0, m - p_1) \\ V_2(p_2, m) &= U(0, 1, m - p_2) \end{aligned}$$

Consequently, the consumer will choose alternative 1 if and only if

$$V_1(p_1, m) \geq V_2(p_2, m)$$

Idiosyncratic Preferences

- In many settings, it is desirable to have a model in which consumers who look identical make different decisions.
- For example, we often observe two consumers with the same income and the same observe characteristics buying different cars.
- For example, one may buy a Honda Accord and the other one buys a Toyota Camry, these are very similar, but not identical choices.
- The differences in the choices must be due to differences in tastes or preferences which are purely idiosyncratic.

Random Utility Models

- We can capture these idiosyncratic taste differences by adding random shocks to the conditional indirect utility function.
- Let ϵ_j thus denote the random shock associated with the conditional indirect utility of product j.
- If these shocks are additively separable to the conditional indirect utility function, then we obtain the following specification of the discrete choice problem:

$$\max_{j=\{1,2\}} \left[V_j(p_j, m) \ + \ \epsilon_j\right]$$

- Utility functions that arise when we add random preference shocks are called random utility functions.

Optimal Choices Revised

- Let's reconsider a simple example with two products and assume that the conditional indirect utility functions are constant (i.e., they do not depend on income or prices).
- Here, a consumer will choose alternative 1 over alternative 2 if and only if:

$$V_1 + \epsilon_1 \geq V_2 + \epsilon_2$$

which we can rewrite as

$$\epsilon_1 - \epsilon_2 \geq V_2 - V_1$$

- So the difference in the idiosyncratic random preference shocks $\epsilon_1 - \epsilon_2$ must larger than the difference in the common components $V_2 - V_1$.

Market Shares

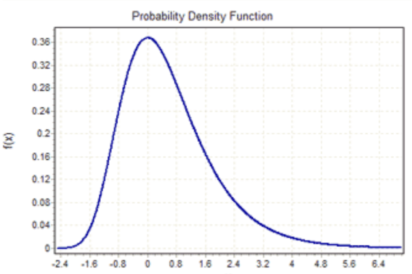
- Suppose we now want to compute the market share of product 1. Let's assume that there are a large number of consumers.
- The market share of product 1 is given by conditional choice probability:

$$s_1 = Pr\{\epsilon_1 - \epsilon_2 \geq V_2 - V_1\}$$

- The conditional choice probability is the probability that an individual will choose product 1.
- It depends on the distribution of the idiosyncratic preference shocks ϵ_1 and ϵ_2. It also depends on the magnitude of the common components, V_1 and V_2.

Type I Extreme Value Distribution

- Following McFadden (1974) we assume that ϵ_j are Type I extreme value errors. The distribution function is illustrated in Figure 1.



- This distribution may look strange, but it is not that different than a normal distribution.

Market Shares with Type I Extreme Value Distribution

- With this functional form, the market share of product 1 has a closed form solution and is given by:

$$s_1 = \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)}$$

- Hence the market share of product 1 increases in V_1 and decreases in V_2.
- However, even when the common utility of 1 (V_1) is much larger than 2 (V_2), some people will still choose 2 because of their idiosyncratic preferences.

An Operational Random Utility Model

- Let's follow Berry (1994) account for unobserved product characteristics. Why do we need those?
- The utility of agent i of product j is given by

$$\begin{aligned} V_{ij} &= x_j'\beta + \alpha p_j + \xi_j + \epsilon_{ij} \\ &= \delta_j + \epsilon_{ij} \end{aligned}$$

  where
  - x_j: observed characteristics,
  - ξ_j: unobserved characteristics,
  - p_j: price,
  - $\delta_j (= V_j)$: mean utility of product j.
- Note that α is negative.

Logit Model Notes

- There is an *unobserved* (to the econometrician) product characteristic (or demand shock), ξ_j.
- This is very important for two reasons:
  1. It will be our econometric error in any regression/estimation.
  2. It is the source of our endogeneity problem.
- Why the endogeneity?
- Because it is assumed to be observed by both consumers and firms
- Products that have high ξ_j are more highly valued by consumers.
- Firms recognize this and price them higher.

Some standard normalizations

When dealing with discrete-choice models, there are two standard normalizations that one must make:

1. Setting the mean utility of one of the goods to zero
   - Note all choices only depend on the *differences* in utility.
   - Equivalently, if we added K "utils" to each product, no choices would change.
   - Thus we set one utility to zero and measure all other utility *relative* to this baseline.
   - Which j to choose? Usually the "outside good," i.e. $\delta_0 = 0$.
2. We can normalize the variance of the random shocks, $\sigma^2_{\varepsilon_j}$.
   - We do this because all we could in principle see is whether someone bought a product (and not how high is her utility).
   - Equivalently, if we multiplied ε_{ij} and every utility parameter, α, β, by K, the choices would not change.
   - Thus we "set the scale of utility" by normalizing the variance of ε_j.

A Feasible MLE

However,

$$\delta = \text{argmax} \;\; L = \Pi_{i=1}^{N} \Pi_{j=0}^{J} \left[\frac{exp\{\delta_j\}}{\sum_{j=0}^{J} exp\{\delta_j\}} \right]^{y_{ij}}$$

is a feasible estimator. Moreover

$$\delta_j = x_j' \beta + \alpha p_j + \xi_j$$

can be interpreted as a linear regression model.

Data on Market Shares and Product Characteristics

The conditional choice probabilities are given by:

$$\begin{aligned}
Pr\{d_{ij} = 1 | x_j, p_j\} &= \frac{exp\{x_j'\beta + \alpha p_j + \xi_j\}}{\sum_{j=0}^{J} exp\{x_j'\beta + \alpha p_j + \xi_j\}} \\
&= \frac{exp\{\delta_j\}}{\sum_{j=0}^{J} exp\{\delta_j\}} \\
&= s_j(\delta)
\end{aligned}$$

which are equal to the aggregate market shares. Why?

Estimation

- Suppose we only have access to "aggregate" data, that is, we only observe $\{s_j, x_j, p_j\}_{j=0}^{J}$.
- If we normalize the mean utility of the outside good to be zero, that is $\delta_0 = 0$, then solving the equation characterizing the market shares yields the following regression model:

$$\ln(s_j) - \ln(s_0) = \delta_j = x_j'\beta + \alpha p_j + \xi_j$$

- Hence, we can recover the δ_j's from the observed market shares.
- Note that the ξ_j's are the error terms of the regression model.

Identifying Assumptions

- To estimate this demand model we need one more assumption.
- $E[\xi_j | x_j, p_j] = 0$ is not a plausible restriction since p_j is likely to be correlated with ξ_j.
- Optimal pricing of firms will imply a correlation between prices and unobserved product characteristics.
- If there exists an instrument z_j such that $E[\xi_j | x_j, z_j] = 0$, we can then estimate the parameters of the model using an IV estimator.
- What are good instruments? Cost shifters, characteristics of products of close competitors, etc.

Some Comments on Aggregation

- To use aggregate data, we need to solve an aggregation problem.
- We start with a model of individual demand and derive the aggregate market shares by aggregating the individual conditional choice probabilities.
- In this simple version of the model, the market share is equal to the conditional choice probability since there is no individual heterogeneity besides the idiosyncratic errors.

TABLE III

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)	
R^2	0.387	*n.a.*	.656

Notes: The standard errors are reported in parentheses.

*The continuous product characteristics—hp/weight, size, and fuel efficiency (*MP$* or *MPG*)—enter the demand equations in levels, but enter the column 3 price regression in natural logs.

Logit Elasticities

- Since I have couched the discussion of demand estimation in terms of elasticities, what are the Logit
- Own-price derivatives / elasticities:

$$\frac{\partial s_j(p)}{\partial p_j} = \frac{\partial s_j(p)}{\partial \delta_j}\frac{\partial \delta_j}{\partial p_j} = s_j(1-s_j)\alpha$$

$$\boxed{\Rightarrow \eta_{jj} = \frac{\partial s_j(p)}{\partial p_j}\frac{p_j}{s_j} = \alpha p_j(1-s_j)}$$

- Cross-price derivatives / elasticities:

$$\frac{\partial s_j(p)}{\partial p_k} = \frac{\partial s_j(p)}{\partial \delta_k}\frac{\partial \delta_k}{\partial p_k} = -s_j s_k \alpha$$

$$\boxed{\Rightarrow \eta_{jk} = \frac{\partial s_j(p)}{\partial p_k}\frac{p_k}{s_j} = -\alpha p_k s_k}$$

Implications of Logit Elasticities

- Note that the cross-price elasticity of good j with respect to good k *is independent of* j!
- This means that the proportionate change in demand for any two products with respect to good k is the same.
  - For example, consider a Mercedes SL and a Honda Civic.
  - Then the cross-price elasticity of each with respect to a change in the price of a Mini Cooper *is the same*, ...
  - ... even though the Mini is "more like" the Civic.

A Final Thought about the Logit Model

- I have said we will consider alternative parameterizations of own- and cross-price elasticities that reduce the number of free parameters.
- Question: How many free parameters does a logit demand model have? Let's look at the elasticity formulas:

$$\eta_{jj} = \alpha p_j (1 - s_j)$$
$$\eta_{jk} = -\alpha p_k s_k$$

- Answer: One free parameter. This is *not* very flexible!
- Fortunately, we can generalize the logit model by allowing for more unobserved heterogeneity (random coefficients) among consumers.
- These generalized logit models then generate more realistic demand patterns.

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

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 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

Applications of Differentiated Product Models in Urban Economics

Differentialed product models are also very useful to estimate spatial sorting models:

- Sorting within a city or a local labor market:
  Epple and Sieg (1999), Sieg, Smith, Banhzaf and Walsh (2004), Bayer, Ferreira and McMillan (2007), Ferreyra (2008), Ahlfelds, Redding, Sturm and Wolf (2015), Epple, Jha, and Sieg (2018), ...
- Sorting across cities and local labor markets:
  Rosen (1979) / Roback (1982), Kennan and Walker (2011), Diamond (2016), ...

Applying BLP to Housing Markets

- You need to define the choice set of the individuals: house, neighborhood, communities
- If you define the choice set not at the house level, you need to estimate neighborhood specific housing prices (Sieg, Smith, Banhzaf and Walsh, 2004).
- You cannot treat each house as a different product, because you need the number of products to go to infinity at a much slower pace than the number of individuals in the sample (Berry, Linton, and Pakes, 2004).
- If you use a logit model then you only need to observe a random subsample of the houses in the choice set because of the IIA property of the logit model (McFadden, 1978).
- For a more careful discussion see the technical appendix of BFM (2007).

BFM Framework

- BMR apply the micro BLP model to estimate household preferences for school and neighborhood attributes in the presence of sorting.
- The model is estimated using restricted access micro-level Census data from the SF metropolitan area using the subsample of houses that within 0.2 miles of a school district boundary.
- They develop an instrument for price that is based on the exogenous attributes of houses and neighborhoods that are located more than 3 miles away from a given house, while allowing the attributes of houses and neighborhoods within 3 miles of the house to directly affect utility.
- They also include boundary fixed effects to account for potential correlation between school quality, neighborhood characteristics and unobserved characteristics In that sense they embed a boundary discontinuity design (Black, 1999) into the model.

BFM Findings

- Households are willing to pay less than 1 percent more in house prices—substantially lower than previous estimates—when the average performance of the local school increases by 5 percent.
- Much of the apparent willingness to pay for more educated and wealthier neighbors is explained by the correlation of these sociodemographic measures with unobserved neighborhood quality.
- Neighborhood race is not capitalized directly into housing prices; instead, the negative correlation of neighborhood percent black and housing prices is due entirely to the fact that blacks live in neighborhoods with unobserved lower-quality.
- Finally, there is considerable heterogeneity in preferences for schools and neighbors, with households preferring to self-segregate on the basis of both race and education.