Demand in Differentiated-Product Markets (part 2)*

Spring 2009

1 Berry (1994): Estimating discrete-choice models of product differentiation

Methodology for estimating differentiated-product discrete-choice demand models, using aggregate data. Fundamental problem is price endogeneity.

Consider the data structure: cross-section of market shares:

j	s_j	p_j	X_1	X_2
A	25%	$1.50	red	large
B	30%	$2.00	blue	small
C	45%	$2.50	green	large

where the total market size is M and there are J brands

Note, in a differentiated goods model where products are valued based on their characteristics, often run into the case where brands with highly-desired characteristics (i.e. higher quality) command higher prices. If the econometrician doesn't observed one (or more) of these characteristics (e.g. quality), the model might find a positive coefficient on price, implying that people, everything else equal, prefer more expensive products!

*These notes rely those produced by Matthew Shum and Aviv Nevo

Background: Trajtenberg (1989) study of demand for CAT scanners. Disturbing finding: coefficient on price is positive, implying that people prefer more expensive machines!

Possible explanation: quality differentials across products not adequately controlled for. Unobserved quality leads to price endogeneity.

Basic demand side model

- Household i's indirect utility from purchasing product j is

$$U_{ij} = X_j\beta - \alpha p_j + \xi_j + \epsilon_{ij} \tag{1}$$

  where ϵ is an iid error term.

- refer to $\delta_j = X_j\beta - \alpha p_j + \xi_j$ as the mean utility from brand j, b/c it is common across all households.

- Econometrician does not observe (ξ_j, ϵ_{ij}) (called structural errors), but the household does. ξ are commonly interpreted as "unobserved quality". All else equal, people prefer more ξ.

- because households and firms observe ξ, it is correlated with p_j (and potentially with X_j, and so the source of the endogeneity problem.

- Assume $\epsilon \sim iid$ TIEV across consumers and products

- Let

$$y_i j = \begin{cases} 1 & \text{if } i \text{ chooses brand } j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- as detailed before, the choice probabilities take the MN logit form

$$Pr(y_i j = 1|\beta, \alpha X, \xi) = \frac{\exp(\delta_j)}{\sum_{k=1}^{J} \exp(\delta_k)} \tag{3}$$

  this is also the predicted share function, $\hat{s}_j$. So want to find values of (α, β, ξ) such that $\hat{s}_j$ is as close as possible to s_j

- note: can't use nonlinear least squares:

$$\min_{\alpha,\beta} \sum_{j=1}^{J} (\hat{s}_j - s_j)^2 \tag{4}$$

because don't observe ξ.

- Berry (1994) suggests a clever IV approach.
  - assume there exists instruments Z such that $E[\xi Z] = 0$
  - Use this equality as a moment condition. The sample analogue is

$$\frac{1}{J}\sum_{j=1}^{J} \xi_j Z_j = \frac{1}{J}\sum_{j=1}^{J} \Big(\delta_j - X_j\beta + \alpha p_j\Big) Z_j = 0 \tag{5}$$

- But how do we calculate δ_j? Two step approach

First step

- The J nonlinear equations equating s_j and $\hat{s}_j$ has J unknowns $\delta_1 \ldots, \delta_J$:

$$\begin{aligned} s_1 &= \hat{s}_1(\delta_1(\alpha,\beta,\xi_1),\ldots,\delta_J(\alpha,\beta,\xi_J)) \\ &\vdots \\ s_J &= \hat{s}_J(\delta_1(\alpha,\beta,\xi_1),\ldots,\delta_J(\alpha,\beta,\xi_J)) \end{aligned}$$

- Invert this system of equations to solve for δ as functions of observed market shares.
- Recall, that typically we normalize the relative utilities by the outside good, and so set $\delta_0 = 0$.
- End result, get $del\hat{t}a_j = \delta_j(s_1, \ldots, s_J)$, for $j = 1, \ldots, J$.

Second step

- Recall that

$$\delta_j = X_j\beta - \alpha p_j + \xi_j \tag{6}$$

- using the estimated $\hat{\delta}$ calculate the sample moment condition

$$\frac{1}{J}\sum_{j=1}^{J}\Big(\delta_j - X_j\beta + \alpha p_j\Big)Z_j \tag{7}$$

  and solve for (α, β) which minimizes this expression.

- Because δ_j is linear in (X, p, ξ) then linear IV methods are applicable. For example, we could use 2SLS, where we regress p on Z in the first stage and obtain the fitted values of $\hat{p}$. Then in the second stage we regress $\hat{\delta}$ on X and $\hat{p}$.

Appropriate instruments?

- In the usual demand case: cost shifters. But note these instruments need to vary across brands in a market. Wages, for example, may not work, since they might be cost shifters for all brands.

- BLP (up next) and to some extent Bresnahan's paper advocate using characteristics, specifically competing products characteristics. Intuition: oligopolistic competition makes firm j set p_j as a function of competing cars characteristics (how close competitors are to product j). However, characteristics of competing cars should not affect household's valuation of firm j's car.

- Nevo (2001) and Hausman (1996) also explore/advocate using prices and market shares of product j in other markets as an instrument for price of product j in market t. Only useful where prices and market shares for same products are observed across many markets.

Example: Logit case

One simple case of inversion step is the MNL case:

$$\hat{s}_j(\delta_1, \ldots, \delta_J) = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^{J} \exp(\delta_k)} \tag{8}$$

The system of equations for matching predicted and actual is (after taking logs)

$$\log(s_1) = \delta_1 - \log(1 + \sum_{k=1}^{J} \exp(\delta_k)) \tag{9}$$

$$\vdots \tag{10}$$

$$\log(s_J) = \delta_J - \log(1 + \sum_{k=1}^{J} \exp(\delta_k)) \tag{11}$$

$$\log(s_0) = \delta_0 - \log(1 + \sum_{k=1}^{J} \exp(\delta_k)) \tag{12}$$

which gives us

$$\delta_j = \log(s_j) - \log(s_0) \tag{13}$$

So in the second step, run an IV regression

$$\log(s_j) - \log(s_0) = X_j\beta - \alpha p_j + \xi_j \tag{14}$$

Measuring market power: recovering markups

- From demand estimation, we have the estimated demand functions for product j, $D^j(X, p, \xi)$.
- Need to specify costs of production, $C^j(q_j, w_j, \omega_j)$. where q is total production, w is observed cost components (can be characteristics of brand, e.g. size), and ω are unobserved cost components (another structural error).
- Then profits for brand j are

$$\Pi_j = D^j(X, p, \xi)p_j - C^j(D^j(X, p, \xi), w_j, \omega_j) \tag{15}$$

- For the multiproduct firm k, profits are

$$\tilde{\Pi}_k = \sum_{j \in \mathcal{K}} \Pi_j = \sum_{j \in \mathcal{K}} \Big[D^j(X, p, \xi)p_j - C^j(D^j(X, p, \xi), w_j, \omega_j) \Big] \tag{16}$$

- Note: we assume that there are no economies of scope, so that production is simply additive across brands for a multiproduct firm (although this could be incorporated into model).
- As with Bresnahan, we need to assume a particular model of oligopoly competition. A common assumption is *Bertrand* price competition.
- Under price competition, equilibrium prices are characterized by J equations

$$\frac{d\tilde{\Pi}_k}{dp_j} = 0, \quad \forall j \in \mathcal{K}, \forall k \tag{17}$$

$$D^j + \sum_{l \in \mathcal{K}} \frac{dD^l}{dp_j} \big(p_l - mc_l)\big) = 0 \tag{18}$$

where mc is the marginal cost function

- If we estimate the demand side, then we already know the demand dunctions D^j and the full set of derivatives, $\frac{dD^j}{dp_j}$ can be computed. Using the FOC, we can solve for the J margins, $p_j - mc_j$, a fairly easy task b/c the system of equations is linear.

$$mc = p + \Omega^{-1} D \tag{19}$$

where derivative matrix Ω

$$\Omega_{ij} = \begin{cases} \frac{dD^i}{dp_j} & \text{if models } (i,j) \text{ are produced by the same firm} \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

- markup measures can then be obtained as $\frac{p_j - mc_j}{p_j}$. Hence, we are using "inverse-elasticities" conditions to calculate market power.

Estimating cost function parameters

- If you want to estimate coefficients in the cost function, need a few extra steps. Need to specify cost function.

- Consider simple case of constant marginal cost

$$mc_j = w_j \gamma + \omega_j \tag{21}$$

where γ are parameters to be estimated. So have the following FOC

$$D^j + \sum_{l \in \mathcal{K}} \frac{dD^l}{dp_j}(p_l - mc_l) = 0 \tag{22}$$

- typically estimate this with a two-step approach (analogous to the demand side)

    1. Inversion: the system of best-responses is J equations with J unknowns.
    2. IV estimation: Estimate the regression $mc_j = w_j \gamma + \omega_j$. Allow for the endogeneity of observed cost components by using demand shifters as instruments. Assume you have instruments W_j such that $E(\omega W) = 0$, then find the γ that minimizes the sample analogue $\frac{1}{J}\sum_{j=1}^{J}(mc_j - w_j \gamma) W_j$.

- Of course, you can estimate the demand and supply side jointly, by jointly imposing the two moment conditions

$$E[\xi Z] = 0 \quad E[\omega W] = 0 \tag{23}$$

This is not entirely straightforward, because the marginal costs are themselves functions of the demand parameters. So we need to employ a more complicated "nested" estimation procedure (outlined in BLP 1995).

2 Berry, Levinsohn, and Pakes (1995): Demand estimation using the random-coefficients logit model

- Return to the demand side. Next we discuss the random coefficients logit model, which is the main topic of Berry, Levinsohn, and Pakes (1995) (aka BLP 1995).

- Assume that utility function is:

$$u_{ij} = X_j\beta_i - \alpha_i p_j + \xi_j + \epsilon_{ij} \tag{24}$$

  The difference here is that the slope coefficients (α_i, β_i) are allowed to vary across households i.

- We assume that, across the population of households, (α_i, β_i) are i.i.d. random variables. The most common assumption is that these random variables are jointly normally distributed:

$$(\alpha_i, \beta_i) \sim N\big((\bar{\alpha}, \bar{\beta}), \Gamma\big) \tag{25}$$

  For this reason, α_i and β_i are called "random coefficients." Hence, $(\bar{\alpha}, \bar{\beta}, \Gamma)$ are additional parameters to be estimated.

- Given these assumptions, the mean utility γ_j is $X_j\bar{\beta} - \bar{\alpha}p_j + \xi_j$ and

$$u_{ij} = \delta_j + \epsilon_{ij} + (\beta_i - \bar{\beta})X_j - (\alpha_i - \bar{\alpha})p_j \tag{26}$$

  so that, even if the ϵ_{ij}'s are still i.i.d. TIEV, the composite error is not. Here, the simple MNL inversion method will not work.

- The estimation methodology for this case is developed in BLP 1995.

- First note: for a given (α_i, β_i), the choice probabilities for household i take the MNL form:

$$Pr(i,j) = \frac{\exp(X_j\beta_i - \alpha_i p_j + \xi_j)}{1 + \sum_{k=1}^{J} \exp(X_k\beta_i - \alpha_i p_k + \xi_k)}. \tag{27}$$

- In the whole population, the aggregate market share is just

$$
\begin{aligned}
\tilde{s}_j &= \int\int Pr(i,j)dG(\alpha_i,\beta_i) && (28)\\
&= \int\int dG(\alpha_i,\beta_i) && (29)\\
&= \int\int \frac{\exp(X_j\beta_i - \alpha_i p_j + \xi_j)}{1+\sum_{k=1}^{J}\exp(X_k\beta_i - \alpha_i p_k + \xi_k)} dG(\alpha_i,\beta_i) && (30)\\
&= \int\int \frac{\exp(\delta_j + \epsilon_{ij} + (\beta_i - \bar{\beta})X_j - (\alpha_i - \bar{\alpha})p_j)}{1+\sum_{k=1}^{J}\exp(\delta_k + \epsilon_{ik} + (\beta_i - \bar{\beta})X_k - (\alpha_i - \bar{\alpha})p_k)} dG(\alpha_i,\beta_i) && (31)\\
&\equiv \tilde{s}_j(\delta_1,\ldots,\delta_J;\bar{\alpha},\bar{\beta},\Gamma) && (32)
\end{aligned}
$$

  that is, roughly speaking, the weighted sum (where the weights are given by the probability distribution of (α,β)) of $Pr(i;j)$ across all households.

- The last equation in the display above makes explicit that the predicted market share is not only a function of the mean utilities $\{\delta_1,\ldots,\delta_J\}$ (as before), but also functions of the parameters $(\bar{\alpha},\bar{\beta},\Gamma)$. Hence, the inversion step described before will not work, because the J equations matching observed to predicted shares have more than J unknowns (i.e. $\delta_1,\ldots,\delta_J,\bar{\alpha},\bar{\beta},\Gamma$).

- Moreover, the last expression is difficult to compute, because it is a multidimensional integral. BLP (1995) propose simulation methods to compute this integral. We will discuss simulation methods later. For the rest of these notes, we assume that we can compute $\tilde{s}_j$ given a vector of parameters

- We would like to proceed, as before, to estimate via GMM, exploiting the population moment restriction $E(\xi Z)=0$. We would like to estimate the parameters by minimizing the sample analogue of the moment condition:

$$
\min_{\bar{\alpha},\bar{\beta},\Gamma} \frac{1}{J}\sum_{j=1}^{J}\big(\delta_j - X_j\bar{\beta} + \bar{\alpha}p_j\big)Z_j \equiv Q(\bar{\alpha},\bar{\beta},\Gamma) \quad (33)
$$

  But problem is that we cannot perform inversion step as before, so that we cannot easily derive $(\delta_1,\ldots,\delta_J)$.

- So BLP propose a "nested" estimation algorithm, with an "inner loop" nested within an "outer loop"
  - In the outer loop, we iterate over different values of the parameters $\theta = \{\bar{\alpha}, \bar{\beta}, \Gamma\}$. Let $\hat{\theta}$ be the current values of the parameters being considered.
  - In the inner loop, for the given parameter values $\hat{\theta}$, we wish to evaluate the objective function $Q(\hat{\theta})$. In order to do this we must:
    1. At current θ, we find the mean utilities $(\delta_1, \ldots, \delta_J)$ which solve the system of equations

$$
\begin{aligned}
s_1 &= \tilde{s}_1((\delta_1, \ldots, \delta_J); \hat{\theta}) \\
&\vdots \\
s_J &= \tilde{s}_J((\delta_1, \ldots, \delta_J); \hat{\theta})
\end{aligned}
$$

       Note that, since we take $\hat{\theta}$ as given, this system is J equations with J unknowns $(\delta_1, \ldots, \delta_J)$.
    2. For the resulting $(\delta_1, \ldots, \delta_J)$, calculate

$$Q(\hat{\theta}) \equiv \frac{1}{J}\big(\delta_j(\hat{\theta}) - X_j \beta_{\hat{\theta}} + \alpha_{\hat{\theta}} p_j\big) Z_j \tag{34}$$

    3. Then we return to the outer loop, which searches until it finds parameter values which minimize $Q(\hat{\theta})$.
  - Essentially, the original inversion step is now nested inside of the estimation routine.
- Within this nested estimation procedure, we can also add a supply side to the BLP model. With both demand and supply-side moment conditions, the objective function becomes:

$$Q(\theta, \gamma) = G_J(\theta, \gamma)' W_J G_j(\theta, \gamma) \tag{35}$$

where G_J is the (M + N)-dimensional vector of stacked sample moment conditions:

$$G_J(\theta_\gamma) = \begin{bmatrix} \frac{1}{J}\sum_{j=1}^J \big(\delta_j(\theta) - X_j\hat{\beta} + \hat{\alpha}p_j\big) z_{1j} \\ \vdots \\ \frac{1}{J}\sum_{j=1}^J \big(\delta_j(\theta) - X_j\hat{\beta} + \hat{\alpha}p_j\big) z_{1j} \\ \frac{1}{J}\sum_{j=1}^J \big(c_j(\theta) - w_j\gamma\big) u_{1j} \\ \vdots \\ \frac{1}{J}\sum_{j=1}^J \big(c_j(\theta) - w_j\gamma\big) u_{Nj} \end{bmatrix}$$

where M is the number of demand side IV's, and N the number of supply-side IV's. W_J is a (M+N)-dimensional weighting matrix. (Assuming M+N $\geq$ dim(θ)+dim(γ))

- The only change in the estimation routine described in the previous section is that the inner loop is more complicated
- In the inner loop, for the given parameter values $(\hat{\theta}, \hat{\gamma})$, we wish to evaluate the objective function $Q(\hat{\theta}, \hat{\gamma})$. In order to do this we must:
    1. At current $\hat{\theta}$, we solve for the mean utilities $\delta_1(\hat{\theta}), \ldots, \delta_J(\hat{\theta})$ as previously.
    2. For the resulting $\delta_1(\hat{\theta}), \ldots, \delta_J(\hat{\theta})$ calculate
    $$\tilde{s}(\hat{\theta}) \equiv \big(\tilde{s}_1(\delta(\hat{\theta})), \ldots, \tilde{s}_J(\delta(\hat{\theta}))\big) \tag{36}$$
    and also the partial derivative matrix
    $$\mathbf{D}(\hat{\theta}) = \begin{bmatrix} \frac{d\tilde{s}_1(\delta(\hat{\theta}))}{dp_1} & \frac{d\tilde{s}_1(\delta(\hat{\theta}))}{dp_2} & \cdots & \frac{d\tilde{s}_1(\delta(\hat{\theta}))}{dp_J} \\ \frac{d\tilde{s}_2(\delta(\hat{\theta}))}{dp_1} & \frac{d\tilde{s}_2(\delta(\hat{\theta}))}{dp_2} & \cdots & \frac{d\tilde{s}_2(\delta(\hat{\theta}))}{dp_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d\tilde{s}_J(\delta(\hat{\theta}))}{dp_1} & \frac{d\tilde{s}_J(\delta(\hat{\theta}))}{dp_2} & \cdots & \frac{d\tilde{s}_J(\delta(\hat{\theta}))}{dp_J} \end{bmatrix}$$
    These derivatives are straightforward to calculate under the MNL case.
    3. Use the supply-side best response equations to solve for $c_1(\hat{\theta}), \ldots, c_J(\hat{\theta})$:
    $$\tilde{s}(\hat{\theta}) + \mathbf{D}(\hat{\theta}) * \begin{pmatrix} p_1 - c_1 \\ \cdots \\ p_J - c_J \end{pmatrix} = 0. \tag{37}$$
    4. So now you can compute $G(\hat{\theta}, \hat{\gamma})$.

3 Computing integrals through simulation

- The principle of simulation: approximate an expectation as a sample average. Validity is ensured by law of large numbers.
- In the case of equation 28, note that the integral there is an expectation:

$$\mathcal{E}(\bar{\alpha}, \bar{\beta}, \Gamma) = E\Big[\frac{\exp(\delta_j + (\beta_i - \bar{\beta})X_j - (\alpha_i - \bar{\alpha})p_j}{1 + \sum_{k=1}^{J} \exp(\delta_k + (\beta_i - \bar{\beta})X_k - (\alpha_i - \bar{\alpha})p_k} | \bar{\alpha}, \bar{\beta}, \Gamma\Big] \tag{38}$$

  where the random variables are α_i and β_i, which we assume to be drawn from the multivariate normal distribution $N((\bar{\alpha}, \bar{\beta}), \Gamma)$.

- For $s = 1, \ldots S$ simulation draws:

  1. Draw (ω_1^s, ω_2^s) independently from $N(0, 1)$
  2. For the current parameter estimates $(\hat{\alpha}, \hat{\beta}, \hat{\Gamma})$, transform (ω_1^s, ω_2^s) into a draw from $N((\hat{\alpha}, \hat{\beta}), \hat{\Gamma})$ using the transformation

$$\begin{pmatrix} \alpha_s \\ \beta_s \end{pmatrix} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} + \hat{\Gamma}^{\frac{1}{2}} \begin{pmatrix} \omega_1^s \\ \omega_2^s \end{pmatrix} \tag{39}$$

     where $\hat{\Gamma}^{\frac{1}{2}}$ is shorthand for the "Cholesky factorization" of the matrix $\hat{\Gamma}$. The Cholesky factorization of a square symmetric matrix Γ is the triangular matrix G such that $G'G = \Gamma$, so roughly it can be thought of a matrix-analogue of "square root". We use the lower triangular version of $\hat{\Gamma}^{\frac{1}{2}}$.

  3. Then approximate the integral by the sample average (over all the simulation draws)

$$\mathcal{E}(\bar{\alpha}, \bar{\beta}, \Gamma) \approx \frac{1}{S} \sum_{s=1}^{S} \Big[\frac{\exp(\delta_j + (\beta^s - \bar{\beta})X_j - (\alpha^s - \bar{\alpha})p_j}{1 + \sum_{k=1}^{J} \exp(\delta_k + (\beta^s - \bar{\beta})X_k - (\alpha^s - \bar{\alpha})p_k}\Big] \tag{40}$$

     For given $(\hat{\bar{\alpha}}, \hat{\bar{\beta}}, \hat{\Gamma})$ the law of large numbers ensure that this approximation is accurate as $S \to \infty$.