Demand Estimation: Random Coefficients Models

Tanja Saxell, Ari Hyytinen, Otto Toivanen & Iivo Vehviläinen

Practicalities

- Exercises, slides, reading list, and zoom links on MyCourses (https://mycourses.aalto.fi/course/view.php?id=34016)
- Return your exercise answers to MyCourses as a pdf or html and include the code
- Exercise sessions: 10.3.-7.4 on Thursdays at 16.15-17.45 in Economicum (seminar room 1, except on 31.3. in seminar room 2)
- Questions about practicalities or exercises? Email to Tuomas Markkula (tuomas.markkula@aalto.fi) or Jaakko Markkanen (jaakko.m.markkanen@aalto.fi)
- My email: tanja.saxell@vatt.fi

Outline: Previous Lectures (Toivanen)

- Discrete choice models
  - Logit
  - Nested logit
- These models are intuitive and easy to implement.
- Why not enough?

Outline: Lectures 3-4 (Saxell)

- Reminder: problems with logit/nested logit
- Random coefficients logit - consumer heterogeneity in preferences, rich substitution patterns
  - Estimator
  - Algorithm
  - Aggregate data
  - Challenges and extensions (micro data)

Outline: Lectures 3-4 (Saxell)

- Reminder: problems with logit/nested logit
- Random coefficients logit - consumer heterogeneity in preferences, rich substitution patterns
  - Estimator
  - Algorithm
  - Aggregate data
  - Challenges and extensions (micro data)

Literature: Lectures 3-4 (Saxell)

- Key paper:
  - BLP = Berry, Steven, Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. Econometrica, 63(4), 841–890.
- For practitioners:
  - Nevo, Aviv. 2000. "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," Journal of Economics & Management Strategy, 9(4), 513–548.
  - Conlon, C. and J. Gortmaker. 2020. "Best Practices for Differentiated Products Demand Estimation with pyblp," The RAND Journal of Economics, *forthcoming*.

Outline: Later Lectures (5-8)

- Lectures 5-6 (Hyytinen): identification and instruments
- Lectures 7-8 (Toivanen): supply side

Problems with Logit

- Independence of irrelevant alternatives (IIA)
  - IIA states that the probability of choosing one product over another do not depend on the presence or absence of other "irrelevant" alternatives.
  - In other words, whatever else is on offer does not matter in the choice between j and m.
  - Implication: when the probability of choosing a given alternative changes, all other choice probabilities change in proportion.

Problems with Logit

- Often unrealistic price elasticities:

$$e_{jk} = \begin{cases} -\alpha p_j(1 - s_j(p, x, y, z)) & \text{if } k = j \\ \alpha p_k s_k(p_k, x_k, y, z) & \text{if } k \neq j. \end{cases} \tag{1}$$

- The own price elasticity e_{jj} is increasing in price (absolute value). Why this is unrealistic - given an example!

Problems with Logit

- Often unrealistic price elasticities:

$$e_{jk} = \begin{cases} -\alpha p_j (1 - s_j(p, x, y, z)) & \text{if } k = j \\ \alpha p_k s_k(p_k, x_k, y, z) & \text{if } k \neq j. \end{cases} \tag{1}$$

- The own price elasticity e_{jj} is increasing in price (absolute value). Why this is unrealistic - given an example!
  - We would think people who buy expensive products are less sensitive to price.
- The cross-elasticity e_{jk}, $k \neq j$ depends only on market share and price of k (but not j!) but not on similarities between goods (IIA).

Source of the Problem

- Source of the problem: no correlation in the preference shock across products.
  - E.g., when the preference shock to BMW is high, the preference shock to Mercedes Benz should also be high, while the preference shock to Fiat should be relatively independent.
- Therefore, the preference shocks between two alternative should be more correlated when they are closer in the characteristics space.
- Most of the extensions try to correct for the above.

Why Important?

The main reason to estimate demand is to quantify demand parameters/elasticities:

- determine responses to and welfare effects of price changes
- determine responses to counterfactual policies (mergers, entry, tax changes etc.)
- used with a supply model to infer markups and market power

Solutions

- Nested logit: assume a particular correlation structure among the structural errors e_{ij}. Within a nest, alternatives are "closer substitutes" than across-nest alternatives.
- Extensions: multi-level tree structure.
- One big problem with nested-logit: need to a-priori group products to nests, this is not trivial (examples?).

Solutions

- Nested logit: assume a particular correlation structure among the structural errors e_{ij}. Within a nest, alternatives are "closer substitutes" than across-nest alternatives.
- Extensions: multi-level tree structure.
- One big problem with nested-logit: need to a-priori group products to nests, this is not trivial (examples?).
  - E.g. housing choice: Level 1: Location (Neighborhood), level 2: Housing Type (Rent, Buy, House, Apt); and Level 3: Housing (Bedrooms)?
  - Or some other combination of these?
- Different nest structures can produce very different result.
- The random coefficients models will try to solve this and provide more general treatment.

Random coefficients model: Berry, Levisohn and Pakes or BLP (1995)

BLP

- Workhorse empirical model of demand (and supply) of differentiated products.
- Many of the ideas in Berry (1994), mostly for simpler models (without random coefs.).
- Many extensions and variations.

BLP

- Random coefficients with individuals heterogeneity $\rightarrow$ rich substitution patterns.
- Requires only aggregate (product and market) level data.
  - Because we can construct the aggregated data from individual level data, all the arguments should go through with the individual choice level data.

BLP

- Explicit about unobservables (to the econometrician), including the nature of endogeneity problem.
  - For example, the econometrician may not observe brand values that are created by advertisement and perceived by consumers.
  - Such unobserved product characteristics is likely to be correlated with the price.
- Use the model to reveal appropriate instruments (based on market competition).
- Propose an algorithm for consistent estimation of the model and standard errors.

Demand Model

- A consumer chooses one of the available options (unit demand).
- There are J differentiated products or inside goods (e.g., different types of cars) $j = 1, .., J$.
- One options should be the outside good, ($j = 0$) i.e. none of the products above (e.g., do not buy a car).
- Note that the model is fairly general, a single option could also be a product bundle, e.g. milk+cheese, shirt+jeans...

Demand Model

- Specification for the conditional indirect utility of consumer i for product j in market t:

$$u_{ijt} = x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + e_{ijt}. \quad (2)$$

  - x_{jt}, p_{jt}: observable product/market characteristics
  - ξ_{jt} unobserved product/market characteristic (demand shock e.g. brand/quality, structural errors on the demand side).
  - e_{ijt} idiosyncratic taste for the product.
- Only utility differences matter, so need a normalization for the outside good, e.g. $u_{i0t} = e_{i0t}$.

Outside Good and Income Effect

- Could also also the utilities to depend on income y_i
- How the preference for the outside good is modeled determines how the individual income affects the choice
- For example, assume (Nevo, 2000)

$$u_{ijt} = x_{jt}\beta_{it} + \alpha(y_i - p_{jt}) + \xi_{jt} + e_{ijt} \tag{3}$$

$$u_{i0t} = \alpha y_i + \xi_{0t} + e_{ijt} \tag{4}$$

- The income level does not affect the choice because the term is common and constant across choices (there is no income effect)
- This is in contrast to the case where we have $\alpha \ln(y_i - p_j)$ instead of $\alpha(y_i - p_j)$ in (3) and $\alpha \ln(y_i)$ instead of αy_i in (4) as in BLP
- For simplicity, assume that there is no income effect

Heterogeneity in Preferences

Utility specification: $u_{ijt} = x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + e_{ijt}$

- Random coefficients: $\beta_{it} = \beta + \sigma v_{it}$
  - For simplicity, just for x_{jt} but could be also for p_{jt}
- e_{ijt}, v_{it} iid across consumers and markets, often:
  - e_{ijt}: iid type 1 extreme value distributed (logit)
  - v_{it}: $N(0, 1)$ or drawn from the distribution of demographics (e.g. income) in market t (mean and std observed in aggregate data)

Random Coefficients

- Products differ in different ways, consumers have heterogeneous preferences over these differences.
- For example, consumers with strong taste for one electricity car will probably like other electricity cars too.
- Random coefficients on product characteristics can capture this.
- Large β_{it}^{k}, strong taste for characteristic x^k
- Consumer i's first (and also second) choice have high values of x^k.
- Key issue as a reminder: produces more sensible substitution patterns.
  - As a result, the degree of competition depends on the degree to which similar products are available.

Data (BLP)

- The data set includes information on (essentially) all car models marketed during the 20 year period beginning in 1971 and ending in 1990.
- Unbalanced panel: car models both appear and exit over this period.
- Identify retail list prices (transaction prices are not easy to find) and other product characteristics.
- Distinguishes which firms produce which model
  - Crucial for the supply model and also for the IVs
- In total, N=2217 model/year observations.

Available Products (BLP)

TABLE 1

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.

Substitution to Outside Good (BLP)

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

Exogenous and Endogenous Characteristics

Utility specification: $u_{ijt} = x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + e_{ijt}$

- exogenous product characteristics x_{jt} (uncorrelated with ξ_{jt})
- endogenous product characteristics p_{jt}, usually the price
  - firms know ξ_{jt} when setting prices.
  - each price depends on $\xi_t = (\xi_{1t}, ..., \xi_{Jt})$.
  - need instruments
- But note that we dot estimate the equation above, utilities are not observable.
- Observed prices and quantities/market shares are both endogenous (simultaneously determined).

Utility Specification: Mean Utility

- Redefine the utility specification:

$$
\begin{aligned}
u_{ijt} &= x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + e_{ijt} && (5)\\
&= \delta_{jt}(x_{jt}, p_{jt}, \xi_{jt}, \theta_1) + u_{ijt}(x_{jt}, v_{it}, \theta_2) + e_{ijt}, && (6)
\end{aligned}
$$

where

- $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ is the mean utility of product j in market t, with $\delta_{0t} = 0$ (normalization)
- a deviation from that mean:

$$
\begin{aligned}
\mu_{ijt} &= u_{ijt} + e_{ijt} && (7)\\
&= x_{jt}\sigma v_{it} + e_{ijt} && (8)\\
&= x_{jt}\tilde{\beta}_i + e_{ijt} && (9)
\end{aligned}
$$

- $\theta_1 = (\beta, \alpha)$, $\theta_2 = \sigma$

Consumer Choice and Market Share

- Consumer i''s choice:

$$a_{it} = \text{arg max}_j u_{ijt}.$$

.

Consumer Choice and Market Share

- Consumer i''s choice:

$$a_{it} = \arg\max_j u_{ijt}.$$

.

- The market share of product j is just an integral over the mass of consumers in the region A_{jt}:

$$\begin{aligned} s_{jt} &= P(a_{it} = j) = \int_{A_{jt}} dF(v, e) \\ &= \int_{A_{jt}} dF_v(v) dF_e(e) \text{ (independence assumption)} \end{aligned}$$

where

$$A_{jt}(\delta_t, x_t, \theta_2) = \{(v_{it}, e_{i0t}, ..., e_{iJt}) : u_{ijt} \geq u_{ikt} \text{ for all } k \in \{0, ..., J\}\}$$

Consumer Choice and Market Share

With Type 1 extreme value distributed error terms (e) and random coefficients, the predicted market share is:

$$s_j(\delta_t, x_t, \theta_2) = \int \frac{\exp[\delta_{jt} + x_{jt}\tilde{\beta}_i]}{1 + \sum_k \exp[\delta_{kt} + x_{kt}\tilde{\beta}_i]} dF_{\tilde{\beta}}(\tilde{\beta}_i|\theta_2)$$

Total Demand for Each Product

- If M_t is a measure of the total number of potential consumers in market t, the total demand for product j is in market t:

$$q_{jt} = M_t \times s_j(\delta_t, x_t, \theta_2) \tag{10}$$

- And for the outside good:

$$q_{0t} = M_t - \sum_{j=1}^{J} M_t \times s_j(\delta_t, x_t, \theta_2) \tag{11}$$

Estimation

Estimation

- Market and product level data (observable): x_t, p_t, s_t, M_t and z_t (instruments).
  - Could also use aggregate data on demographics such as income (later).
- How would you measure M_t and the market share of the outside good $s_{0t} = 1 - \sum_{j=1}^{J} s_{jt}/M_t$?

Estimation

- Market and product level data (observable): x_t, p_t, s_t, M_t and z_t (instruments).
  - Could also use aggregate data on demographics such as income (later).
- How would you measure M_t and the market share of the outside good $s_{0t} = 1 - \sum_{j=1}^{J} s_{jt}/M_t$?
  - BLP: M_t is the number of households in the U.S. taken for each year from the Statistical Abstract of the U.S.
- Assume that θ_2 (and the distributions F_v and F_e) are already known.
- For each market t, find $\delta_t \in R^J$ such that $s_j(\delta_t, x_t, \theta_2) = s_{jt}\ \forall j$.
  - Invert market shares to recover mean utilities δ_t.
  - Done this way, δ_t is such that the predicted market shares fit the observed market shares exactly.

Estimation: To-Do

- Instruments
- Inversion step: from market shares to mean utilities
- Formally, define an estimator and algorithm
- (Add supply model)

Price Endogeneity

- Identification concerns: price endogeneity (correlates with ξ_{jt}).
- Need instruments: variables that exogenously shift prices and quantities independently.

BLP Instruments

- IVs based on market competition.
  - In oligopolistic competition, firm *j* sets the price as a function of characteristics of products produced by competing firms.
  - However, characteristics of competing products should not depend on a consumer's valuation of firm j's product.
  - Similarly, for multiproduct firms, can construct IVs using characteristics of all other products produced by same firm *j*.

BLP Instruments

- The following are used as IVs for the price of product in a given market, p_{jt}

$$\sum_{k \neq j \in \mathcal{J}_t \cap \mathcal{F}_f} x_{kt},$$

$$\sum_{k \neq j \in \mathcal{J}_t \setminus \mathcal{F}_f} x_{kt}.$$

  where f is the firm that owns product j and $\mathcal{F}_f$ is the set of products firm f owns
- For example, if one of the characteristics is the size of a car, then the IVs for product j includes the sum of size across own-firm products and the sum of size across rival firm products

Alternative Instruments

- Traditional cost shifters; however need variation in costs across alternatives
- Proxies of cost shifters; price of the same product in other markets (Hausman instruments), valid if demand shocks are uncorrelated across markets
- Characteristics of nearby markets (Waldfogel instruments, after Waldfogel 2003)
- Exogenous shifters of market structure (e.g., firm ownership) that affect prices through equilibrium markups
- More on instruments and identification later (Hyytinen, lectures 5-6)

Inversion Step

- Find δ_t solves the nonlinear system $s_t = s(\delta_t, x_t, \theta_2)$, or equivalently

$$\delta_t = \delta_t + \underbrace{ln(s_t)}_{Data!} - \underbrace{ln(s(\delta_t, x_t, \theta_2))}_{Model\ prediction!} \tag{12}$$

- They show that under mild conditions on the linear random coef. random utility model, $T(\delta_t) = \delta_t + ln(s_t) - ln(s(\delta_t, x_t, \theta_2))$ is a contraction mapping.
- This means that
  - it has a (unique) fixed point in δ_t.
    - $s_t = s(\delta_t, x_t, \theta_2)$ has an inverse $\delta_t = D^{-1}(s_t, x_t, \theta_2)$.
    - We can therefore perform a non-linear change of variables from observed market shares (s_t), x_t and θ_2 to δ_t (see Berry and Haile, 2014).

Analytical Inversion: Logit

Recall the utility specification:

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \varepsilon_{ijt}, \quad s_{jt} = \frac{\exp[x_{jt}\beta - \alpha p_{jt} + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta - \alpha p_{kt} + \xi_{kt}]}$$

- ξ_{jt} potentially correlated with price $\text{Corr}(\xi_{jt}, p_{jt}) \neq 0$
- But not characteristics $E[\xi_{jt} | x_{jt}] = 0$.

Analytical Inversion: Logit

Taking logs:

$$
\begin{aligned}
\ln(s_{0t}) &= -\ln\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right) \\
\ln(s_{jt}) &= [x_{jt}\beta - \alpha p_{jt} + \xi_{jt}] - \ln\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right) \\
\underbrace{\ln(s_{jt}) - \ln(s_{0t})}_{Data!} &= x_{jt}\beta - \alpha p_{jt} + \xi_{jt}
\end{aligned}
$$

Exploit the fact that we have one ξ_{jt} for every share s_{jt} (one to one mapping)

IV Logit Estimation

1. Transform the data: $\ln(s_{jt}) - \ln(s_{0t})$.
2. IV Regression of: $\ln(s_{jt}) - \ln(s_{0t})$ on $x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ with IV z_{jt}.

Analytical Inversion: Nested Logit (Berry 1994)

For nested logit, the same as logit plus an extra term $\ln(s_{j|g})$ the **within group share**:

$$\underbrace{\ln(s_{jt}) - \ln(s_{0t}) - \sigma \ln(s_{j|gt})}_{\text{Data!}} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

$$\ln(s_{jt}) - \ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + \sigma \ln(s_{j|gt}) + \xi_{jt}$$

- Note that $\ln(s_{j|g})$ is also endogenous – we are regressing Y on a function of Y.
- A common instrument for σ is the number of products within the nest.

Inversion: BLP (Random Coefficients)

We can't solve for δ_{jt} analytically this time.

$$s_j(\delta_t, x_t, \theta_2) = \int \frac{\exp[\delta_{jt} + x_{jt}\tilde{\beta}_i]}{1 + \sum_k \exp[\delta_{kt} + x_{kt}\tilde{\beta}_i]} dF_{\tilde{\beta}}(\tilde{\beta}_i|\theta_2)$$

- This is a $J \times J$ system of equations for each t.
- Model predictions $s_j(\delta_t, x_t, \theta_2)$ involve high-dimensional integrals, use simulation (Monte Carlo Integration) to approximate it ("method of simulated moments" instead of GMM).
- There is a unique vector $\boldsymbol{\delta_t}$ that solves it for each market t.
- We can solve δ_t recursively (because the contraction mapping has a unique fixed point) at each trial value of θ_2 (BLP "nested fixed point algorithm").

BLP Estimator (Without Supply Side)

- GMM estimator of $\theta = (\theta_1, \theta_2)$:

$$\min_{\theta} g(\xi(\theta))' W g(\xi(\theta)) \text{ s.t.}$$

  - $g(\xi(\theta)) = \frac{1}{N} \sum_{j,t} \xi_{jt}(\theta)' z_{jt}$
  - $\xi_{jt}(\theta) = \delta_{jt}(\theta_2) - x_{jt}\beta - \alpha p_{jt}$ where $\delta_{jt}(\theta_2) \equiv \delta_j(s_t, x_t, \theta_2)$
  - $s_{jt} = s_j(\delta_t, x_t, \theta_2)$
  - $s_j(\delta_t, x_t, \theta_2) = \int \frac{\exp[\delta_{jt}(\theta_2) + x_{jt}\tilde{\beta}_i]}{1 + \sum_k \exp[\delta_{kt}(\theta_2) + x_{kt}\tilde{\beta}_i]} dF_{\tilde{\beta}}(\tilde{\beta}_i | \theta_2)$, approximation via simulation
  - W: standard GMM weighting matrix: a consistent estimate of $E(z'\xi\xi'z)^{-1}$

BLP Estimator (Without Supply Side)

- GMM estimator of $\theta = (\theta_1, \theta_2)$:

$$\min_{\theta} g(\xi(\theta))' W g(\xi(\theta)) \text{ s.t.}$$

  - $g(\xi(\theta)) = \frac{1}{N} \sum_{j,t} \xi_{jt}(\theta)' z_{jt}$
  - $\xi_{jt}(\theta) = \delta_{jt}(\theta_2) - x_{jt}\beta - \alpha p_{jt}$ where $\delta_{jt}(\theta_2) \equiv \delta_j(s_t, x_t, \theta_2)$
  - $s_{jt} = s_j(\delta_t, x_t, \theta_2)$
  - $s_j(\delta_t, x_t, \theta_2) = \int \frac{\exp[\delta_{jt}(\theta_2) + x_{jt}\tilde{\beta}_i]}{1 + \sum_k \exp[\delta_{kt}(\theta_2) + x_{kt}\tilde{\beta}_i]} dF_{\tilde{\beta}}(\tilde{\beta}_i | \theta_2)$, approximation via simulation
  - W: standard GMM weighting matrix: a consistent estimate of $E(z'\xi\xi' z)^{-1}$
    - At the true parameter value, θ^*, the moment condition $E(z_t \xi_t(\theta^*)) = 0$
    - The weight matrix defines the metric by which we measure how close to zero we are
    - By using the inverse of the variance-covariance matrix of the moments, we give less weight to those moments that have a higher variance

Contraction: BLP

BLP propose an algorithm to find δ_{jt} s.t. $s_{jt} = s_j(\delta_t, x_t, \theta_2)$. Fix θ_2 and solve for δ_t.

$$\delta_{jt}^{(k)} = \delta_{jt}^{(k-1)} + \left[\underbrace{\ln(s_{jt})}_{Data!} - \underbrace{\ln(s_j(\delta_t^{(k-1)}, x_t, \theta_2))}_{Model\ prediction!}\right]$$

- Idea: begin by evaluating the right-hand side of eq. at some initial guess for vector δ_t^0, obtain a new δ_t^1 as the output of this calculation for all j in market t, substitute it back into the right hand side of eq., and repeat this process until convergence.
- If iterate until $|\delta_t^{(k)} - \delta_t^{(k-1)}| < \epsilon_{tol}$ you can recover the δ's so that the observed shares and the predicted shares are identical.
- ϵ_{tol} has to be small (loose tolerance value can make performance poor).
- $s(\delta_t^{(k-1)}, \theta_2)$ requires computing the numerical integral each time (e.g., via monte carlo, later on this).

BLP Algorithm: Basic Idea

- Outer loop: search over trial values of the parameter vector $\theta = (\theta_1, \theta_2)$
- Inner loop: given θ, find a solution for $\delta_t(\theta_2)$ in each market t such that $s_{jt} = s_j(\delta_t, x_t, \theta)$ as fixed point iteration
- Then calculate $\xi_{jt} \equiv \delta_{jt}(\theta_2) - (x_{jt}\beta - \alpha p_{jt})$

begin outer loop
try new θ
    begin inner loop
      solve contraction mapping (fixed point iteration)
    end inner loop
calculate GMM criterion
end outer loop

BLP Pseudocode

From the outside, in:

- Outer loop: search over parameters $\theta = (\theta_1, \theta_2)$ to minimize GMM objective:

$$\widehat{\theta_{BLP}} = \arg\min_{\theta} g(\xi(\theta))' W g(\xi(\theta)) \tag{13}$$

BLP Pseudocode

From the outside, in:

- Outer loop: search over parameters $\theta = (\theta_1, \theta_2)$ to minimize GMM objective:

$$\widehat{\theta_{BLP}} = \arg\min_{\theta} g(\xi(\theta))' W g(\xi(\theta)) \tag{13}$$

- Inner Loop:
  - Fix a guess of θ_2.
  - Solve for δ_{jt} which satisfies $s_j(\delta_t, x_t, \theta_2) = s_{jt}$.
    - **Simulated moments:** computing $s_j(\delta_t, x_t, \theta_2)$ requires numerical integration (simulation).

BLP Pseudocode

From the outside, in:

- Outer loop: search over parameters $\theta = (\theta_1, \theta_2)$ to minimize GMM objective:

$$\widehat{\theta_{BLP}} = \arg\min_{\theta} g(\xi(\theta))' W g(\xi(\theta)) \tag{13}$$

- Inner Loop:
  - Fix a guess of θ_2.
  - Solve for δ_{jt} which satisfies $s_j(\delta_t, x_t, \theta_2) = s_{jt}$.
    - **Simulated moments:** computing $s_j(\delta_t, x_t, \theta_2)$ requires numerical integration (simulation).
- We can do IV-GMM to recover θ

$$\delta_{jt}(\theta_2) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} \rightarrow \xi_{jt}(\theta_1, \theta_2)$$

- Use $\hat{\boldsymbol{\xi}}(\theta)$ to approximate $g(\xi(\theta)) \approx \frac{1}{JT} \sum_{j,t} Z'_{jt} \xi_{jt}$
- Plug into the GMM objective function and approximate W (13)
- Iterate until convergence
- Standard errors: standard MSM (method of simulated moments)

Linear and Nonlinear Parameters

- Important simplification: θ_1 enter objective function and $\xi_{jt} \equiv \delta_{jt}(\theta_2) - (x_{jt}\beta - \alpha p_{jt})$ linearly
- Given θ_2 and W, we have closed-form expression for optimal θ_1 (as a function of θ_2)
- Outer loop (nonlinear) search only involves θ_2
  - The nonlinear parameters θ_2 solve for the mean utility levels δ_{jt} that set the predicted market shares equal to the observed market shares

Approximating Market Shares: Numerical Integration

- Model predictions $s_j(\delta_t, x_t, \theta_2)$ involve high-dimensional integrals, a common approach is to use Monte Carlo Integration to approximate them
- MC integration is a technique for numerical integration using random numbers
- Particularly useful for higher-dimensional integrals

Numerical Integration: Example

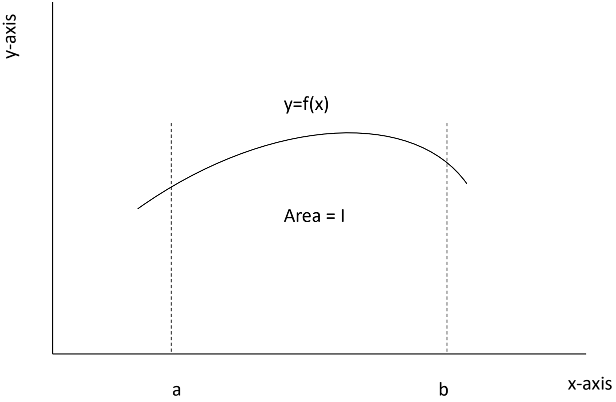

Figure 1: Approximate $I = \int_a^b f(x)dx$

Numerical Integration: Approximate $I = \int_a^b f(x)dx$

- In the simplest (deterministic) approach, the integral is approximated by a summation over N points at a regular interval $\triangle x$ for x:

$$\hat{I} = \sum_{i=1}^{N} f(x_i) \triangle x$$

where $x_i = a + (i - 0.5) \triangle x$ and $\triangle x = \frac{b-a}{N}$, i.e.

$$\hat{I} = \frac{b-a}{N} \sum_{i=1}^{N} f(x_i)$$

- Takes the value of f from the midpoint of each interval

Numerical Integration: Approximate $I = \int_a^b f(x)dx$

- In the simplest (deterministic) approach, the integral is approximated by a summation over N points at a regular interval $\triangle x$ for x:

$$\hat{I} = \sum_{i=1}^{N} f(x_i) \triangle x$$

where $x_i = a + (i - 0.5) \triangle x$ and $\triangle x = \frac{b-a}{N}$, i.e.

$$\hat{I} = \frac{b-a}{N} \sum_{i=1}^{N} f(x_i)$$

- Takes the value of f from the midpoint of each interval
- The sampling method for MC integration is very similar to the simple approach
- Instead of sampling at regular intervals $\triangle x$, we now sample at random points x_i, and then take the average over NS values of these

BLP: Approximating Market Shares

- Approximation of predicted shares, given θ
  $s_j(\delta_t, x_t, \theta_2) = \int \frac{\exp[\delta_{jt}(\theta_2)+x_{jt}\tilde{\beta}_i]}{1+\sum_k \exp[\delta_{kt}(\theta_2)+x_{kt}\tilde{\beta}_i]} dF_{\tilde{\beta}}(\tilde{\beta}_i|\theta_2)$.
- Draw NS values of v_{it} e.g. from $N(0,1)$ to get $\tilde{\beta}_{it} = \sigma v_{it}$.
- Approximate: $s_j(\delta_t, x_t, \theta_2) \approx \frac{1}{NS} \sum_{i=1}^{NS} \frac{\exp[\delta_{jt}(\theta_2)+x_{jt}\tilde{\beta}_{it}]}{1+\sum_k \exp[\delta_{kt}(\theta_2)+x_{kt}\tilde{\beta}_{it}]}$.
- Use the same set of draws for each value θ (outer loop).

Illustrating Benefits of BLP - Reminder of Problems with Logit

- Logit model had the problem of IIA (independence of irrelevant alternatives)
- Under the IIA, the ratio of choice probabilities between two alternatives depend only on the mean utility of these two alternatives and are independent of irrelevant alternatives

$$\frac{s_j(\delta_t)}{s_l(\delta_t)} = \frac{\exp[x_{jt}\beta - \alpha p_{jt} + \xi_{jt}]}{\exp[x_{lt}\beta - \alpha p_{lt} + \xi_{lt}]}$$

Illustrating Benefits of BLP - Reminder of Problems with Logit

- The own price elasticity e_{jj} was increasing in price (absolute value)
- The cross-elasticity e_{jk}, $k \neq j$ depended only on market share and price of k (but not j!) but not on similarities between goods (IIA)

BLP: No IIA

- There is no IIA at the aggregate (market) level:

$$\frac{s_j(\delta_t, x_t, \theta_2)}{s_l(\delta_t, x_t, \theta_2)} = \frac{\int \frac{\exp[\delta_{jt}(\theta_2)+x_{jt}\tilde{\beta}_i]}{1+\sum_k \exp[\delta_{kt}(\theta_2)+x_{kt}\tilde{\beta}_i]} dF_{\tilde{\beta}}(\tilde{\beta}_i|\theta_2)}{\int \frac{\exp[\delta_{lt}(\theta_2)+x_{lt}\tilde{\beta}_i]}{1+\sum_k \exp[\delta_{kt}(\theta_2)+x_{kt}\tilde{\beta}_i]} dF_{\tilde{\beta}}(\tilde{\beta}_i|\theta_2)}$$

- The ratio of market shares depends on the price and characteristics of all the other products

BLP: Price Elasticities

- Finally, using the predicted market shares, the price elasticities are

$$e_{jk} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha s_{ijt}(1 - s_{ijt}) dF_{\tilde{\beta}}(\tilde{\beta}_i|\theta_2) & \text{if } j = k \\ \frac{p_{kt}}{s_{jt}} \int \alpha s_{ijt} s_{ikt} dF_{\tilde{\beta}}(\tilde{\beta}_i|\theta_2) & \text{otherwise,} \end{cases} \tag{14}$$

where $s_{ijt} = \frac{\exp[\delta_{jt}(\theta_2) + x_{jt}\tilde{\beta}_i]}{1 + \sum_k \exp[\delta_{kt}(\theta_2) + x_{kt}\tilde{\beta}_i]}$

BLP: Price Elasticities (Nevo, 2000)

- The price elasticities depend on the density of unobserved consumer types
- 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
- The price elasticity will be different for different brands
- So if, for example, consumers of BMW have low price sensitivity, then the own-price elasticity of BMW will be low despite the high prices

BLP: Price Elasticities (Nevo, 2000)

- Therefore, substitution patterns are not driven by functional form, but by the differences in the price sensitivity
- The full model also allows for flexible substitution patterns, which are not constrained by a priori segmentation of the market (vs nested logit)

Part 1: Summary

- BLP give us both a statistical **estimator** and an **algorithm** to obtain estimates.
- Attractive for many differentiated products markets
- Flexible substitution patterns, addressing endogeneity concerns
- Widely used in IO but also in other fields:
  - Economics is about making choices (demand and supply)
  - Choices differ, usually in some unobservable ways
- Computationally demanding, progress on these challenges
- Many variations of the model in the literature

Part 2: BLP

- Computationally demanding, progress on these challenges
- Many variations of the model in the literature

Part 2: Random Coefficient Models / BLP

Part 2: Random Coefficient Models / BLP

- Computational challenges
- Best practices
- Extensions (micro data etc)

Part 2: Key Papers

- Conlon, C. and J. Gortmaker. 2020. Best Practices for Differentiated Products Demand Estimation with pyblp. The RAND Journal of Economics, *forthcoming*.
- Dubé, J.-P., Fox, J.T. and Su, C.-L. 2012. Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation. Econometrica, 80: 2231-2267.
- Petrin, A. 2002. Quantifying the Benefits of New Products: The Case of the Minivan. Journal of Political Economy, 110(4).

Computational challenges and best practices

BLP Algorithm: Reminder

$$u_{ijt} = \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}} + x_{jt}\sigma v_{it} + e_{ijt}$$

- Outer loop: search over trial values of the parameter vector $\theta = (\theta_1, \theta_2)$, with $\theta_1 = (\beta, \alpha)$ (linear parameters) $\theta_2 = \sigma$ (nonlinear parameters)
- Inner loop: given θ, find a solution for $\delta_t(\theta_2) \equiv \delta_j(s_t, x_t, \theta_2)$ in each market t such that $s_{jt} = s_j(\delta_t, x_t, \theta)$ as fixed point iteration
- Then calculate $\xi_{jt} \equiv \delta_{jt}(\theta_2) - (x_{jt}\beta - \alpha p_{jt})$

```
begin outer loop
try new θ
    begin inner loop
      solve contraction mapping (fixed point iteration)
    end inner loop
calculate GMM criterion
end outer loop
```

BLP Challenges

- The BLP has a unique fixed point, and the fixed-point iteration is guaranteed to converge from any starting point
- However, the BLP contraction mapping $\delta_t^{(k)} = \delta_t^{(k-1)} + \left[\ln(s_t) - \ln(s(\delta_t^{(k-1)}, x_t, \theta_2))\right]$ converges only linearly and can prove to be a time-consuming procedure, especially in applications with a large number of observations
  - If sequence $\{x^k\}$ converges to x^* and there exists $\lambda > 0$ and $q \geq 1$ s.t. $\lim_{n\to\infty} \frac{|x^{k+1} - x^*|}{|x^k - x^*|^q} = \lambda$ we say that q is the rate of convergence
  - q represents how quickly the sequence approaches its limit
  - $q = 1$: linear, $q = 2$: quadratic etc., ordered from slowest to fastest.

BLP: Challenges (e.g. Conlon and Gortmaker, 2020)

- Inner loop can be slow
- Outer loop optimization is hard
- Involves a nonlinear change of variables from the space of observed market shares to the space of mean utilities for products
- Parameters governing the nonlinear change of variables are unknown
- This results in a non-linear, non-convex optimization problem with a simulated objective function
- Optimization routines such as Nelder-Mead (e.g. Matlab's fminsearch) can fail (convergence to a local minimum)
  - Recommendation: consider a number of different starting values and optimization routines.

BLP: Challenges

The main challenge of the BLP nested fixed point algorithm is solving the system of market shares $s_t = s(\delta_t, x_t, \theta_2)$:

$$\delta_t^{(k)} = \delta_t^{(k-1)} + \left[\ln(s_t) - \ln(s(\delta_t^{(k-1)}, x_t, \theta_2))\right]$$

- Although mathematically there is a unique solution, it is impossible, numerically speaking, to choose a vector δ_t that solves $s_{jt} = s(\delta_t, \theta_2)$ exactly.
- Instead, we must solve the system of equations to some tolerance.

BLP: Challenges

- Tolerances are important, with less stringent stopping criterion for the inner loop can get wrong answers (e.g. Dube, Fox, and Su, 2012)
- The inner loop error propagates into the outer loop GMM objective function and its derivatives, which may cause an optimization routine to fail to converge
- It is also possible to set a tolerance which is too tight and thus can never be satisfied (Conlon and Gortmaker, 2020).
  - They prefer to set the tolerance between 1E-14 and 1E-12 as the machine epsilon or detectable difference between two double precision floating point numbers is around 1E-16.

Accelerating Contraction (Conlon and Gortmaker, 2020)

- A direct approach would be to solve a system of equations $s_t = s(\delta_t, x_t, \theta_2)$ using Newton-type methods.
- Newton's method is a powerful technique - in general the convergence is quadratic.
- Idea of Newton's method for solving a system of equations $f(x) = 0$:
  - Using basic calculus, the linear approximation of f at x_0 is $f(x) \approx f(x_0) + D(f(x_0))(x - x_0)$ where $D(f(x_0))$ is a matrix of partial derivatives (**Jacobian matrix**) evaluated at x_0.
  - To get new x_1, we can solve this equation by $x_1 = x_0 - \Psi^{-1}(f(x_0))f(x_0)$.
  - Iterate until convergence, $(x_k - x_{k-1}) < \epsilon_{tol}$.
  - Matlab: fsolve

Accelerating Contraction (Conlon and Gortmaker, 2020)

Conlon and Gortmaker propose the following Newton (also known as Newton-Raphson) iteration for solving $s_t = s(\delta_t, x_t, \theta_2)$ for step size λ

$$\delta_t^{(k)} = \delta_t^{(k-1)} - \lambda \Psi^{-1}(\delta_t^{(k-1)}, x_t, \theta_2) s(\delta_t^{k-1}, x_t, \theta_2)$$

- Each Newton-Raphson iteration would require computation of:
- A vector of market shares $s_t(\cdot)$
- The Jacobian matrix $\Psi_t(\cdot)$ as well as its inverse (can be costly, ways to speed up).
  - Variants of the algorithm generally involve modifying the step-size λ or approximating $\Psi^{-1}(\cdot)$ in ways that avoid calculating the inverse at each step.

BLP: Challenges

- BLP algorithm also forces constraints on market shares and linear parameters hold exactly at every guess of θ_2 when we care only about constraints' holding at the final solution.
- Progress also on this challenge.

Constrained Optimization

- E.g. Dube, Fox, and Su (2012): alternative computation approach for the same BLP estimator.
- They propose a routine to solve a constrained optimization problem: minimize GMM objective function over parameters, subject to constraint that the inner loop fixed point equations hold.
- Single-step optimization, no inner/outer loop.

DFS Algorithm: Basic Idea

Constrained optimization algorithm of BLP estimator:

$$\begin{aligned}
\min_{\theta,\xi} \quad & \xi' z' W z' \xi \quad \text{s.t.} \\
log(s_{jt}) \quad &= \quad log(s_j(x_t, \xi_t, \theta)) \ \forall j, t \\
\text{where } s_j(x_t, \xi_t, \theta) \quad &= \quad \frac{1}{NS} \sum_{i=1}^{NS} \frac{exp(x_{jt}\beta + \xi_{jt} + x_{jt}\tilde{\beta}_{it})}{1 + \sum_k exp(x_{kt}\beta + \xi_{kt} + x_{kt}\tilde{\beta}_{it})}
\end{aligned} \tag{15}$$

Note that here ξ_{jt} are parameters whose values, together with θ, must equate predicted and observed shares.

- A large number of parameters!

DFS and Constrained Optimization

- Could use e.g. Matlab's fmincon or KNITRO which is a solver for large scale nonlinear mathematical optimization problems
- Fined-tuned by engineers for solving general constrained optimization problems
- Disadvantage: in practice, may have to code both 1st and 2nd derivatives (Hessians) to solve the problem efficiently/well enough:
  - Calculating 1st derivative can be quite tedious and error-prone
  - Calculating 2nd derivatives is even more tedious and fraught with opportunities for error

DFS vs. BLP

- DFS: Solving a constrained optimization problem w.r.t. θ and ξ

$$\begin{aligned} \min_{\theta,\xi} \quad & \xi' z' W z' \xi \quad \text{s.t.} \\ log(s_{jt}) \quad = \quad & log(s_j(x_t, \xi_t, \theta)) \ \forall j, t \end{aligned}$$

  The algorithm is a **mathematical program with equilibrium constraints** (MPEC)

- BLP: Nesting a contraction mapping for each trial value of θ

$$\min_{\theta} \quad \xi(\theta)' z' W z' \xi(\theta)$$

  A **nested fixed point** (NFP) algorithm.

DFS vs. BLP

- DFS: Solving a constrained optimization problem w.r.t. θ and ξ

$$\begin{aligned} \min_{\theta,\xi} & \quad \xi' z' W z' \xi \quad \text{s.t.} \\ log(s_{jt}) &= log(s_j(x_t, \xi_t, \theta)) \; \forall j, t \end{aligned}$$

  The algorithm is a **mathematical program with equilibrium constraints** (MPEC)
- BLP: Nesting a contraction mapping for each trial value of θ

$$\min_{\theta} \quad \xi(\theta)' z' W z' \xi(\theta)$$

  A **nested fixed point** (NFP) algorithm.
- By eliminating the nested calls to a contraction mapping, MPEC can be faster than NFP, especially if the 1st and 2nd derivates are provided to the optimizer.

DFS vs. BLP

- By eliminating the nested calls to a contraction mapping, MPEC can be faster than NFP, especially if the 1st and 2nd derivates are provided to the optimizer.
- Contraction mapping converges linearly in (the original) NFP vs. Newton's method (MPEC) converges quadratically
- Concern: MPEC has a large number of auxiliary parameters
- MPEC: trade-off between quadratic convergence and dimension of optimization
  - May perform poorly with large numbers of markets?

DFS (MPEC) vs. BLP (NFP) with a Tight Inner-Loop Tolerance

TABLE V

MONTE CARLO RESULTS VARYING THE NUMBER OF MARKETS, PRODUCTS, AND SIMULATION DRAWS[a]

# 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	

[a]There is one data set and five starting values for each experiment. The mean intercept is 2. MPEC and NFP produce the same lowest objective value. See the footnote to Table III for other details.

DFS vs. BLP

- BLP and DFS define different **algorithms** to produce the same statistical **estimator**.
  - The unknown parameters satisfy the same set of first-order conditions. (Not only asymptotically, but in finite sample).
  - $\hat{\theta}_{NFP} \approx \hat{\theta}_{MPEC}$ but for numerical differences in the optimization routine.
- A choice of algorithm should mostly be about computational convenience, both can work well when carefully executed!
- Current best practice of BLP, with several improvements to the original version: pyBLP (Python) implementation of Conlon and Gortmaker (2019)

Extensions: demographics, panel data, micro data

BLP with Demographics (and Aggregate Data)

- Can also for consumer preferences to vary as a function of the individual characteristics such as income and age (already in the original BLP)
- A few ways to do this:
  - Use cross sectional variation in s_{jt} and $\overline{y}_t$ (mean or median income).
  - Better: Divide up your data into additional "markets" by demographics: do you observe s_{jt} at this level?
  - Better: Draw y_{it} from a geographic specific income distribution. Draw ν_i from a general distribution of unobserved heterogeneity.

BLP with Demographics (and Aggregate Data)

- Example: Nevo (2000): the distribution of consumer taste parameters as a function of demographics d:

$$\beta_{it} = \beta + \Pi d_{it} + \sigma\nu_{it},\ \ d_{it} \sim F_d, \nu_{it} \sim F_\nu$$

- Given that no individual data is observed in BLP/Nevo (2000), neither component of the individual characteristics is directly observed in the choice data set.
- We know something about the distribution of the demographics:
  - a nonparametric distribution estimated from other data sources
  - a parametric distribution with the parameters estimated elsewhere, e.g. when the mean and standard deviation (μ_d and σ_d) are known, one alternative is to assume $d_i \sim N(\mu_d, \sigma_d)$

Market Share with Demographics

- The market share of product j is almost as before:

$$
\begin{aligned}
s_{jt} &= P(a_{it} = j) = \int_{A_{jt}} dF(v, e, d) \\
&= \int_{A_{jt}} dF_v(v) dF_e(e) dF_d(d) \text{ (independence assumption).}
\end{aligned}
$$

- With Type 1 extreme value distributed error terms (e), this can be approximated by

$$
s_j(\delta_t, x_t, \theta_2) \approx \frac{1}{NS} \sum_{i=1}^{NS} \frac{\exp[\delta_{jt}(\theta_2) + x_{jt}(\Pi d_i + \sigma \nu_i)]}{1 + \sum_k \exp[\delta_{kt}(\theta_2) + x_{kt}(\Pi d_i + \sigma \nu_i)]}
$$

where d_i and ν_i and draws from the empirical counterparts of F_v and F_d

Aggregate Panel Data

- With enough observations on the same product it is possible to include fixed effects

$$\delta_{jt}(\widetilde{\theta}_2) = x_{jt}\beta - \alpha p_{jt} + \underbrace{\xi_{jt}}_{\xi_j+\xi_t+\Delta\xi_{jt}}$$

- What does ξ_j mean in this context?
- What would ξ_t mean in this context?
- $\Delta\xi_{jt}$ is now the structural error term, this changes our identification strategy a little.
- We need instruments that change **within product and across market**.
  - ie: $z_{jt} - \bar{z}_{\cdot t} - \bar{z}_{j\cdot} = \Delta z_{jt}$ has to have some variation left!

Different Types of Data

- Aggregate data: market and product characteristics, prices, shares (cross-sectional, panel)
- Micro data: individual characteristics matched to individual consumer choices
- Panel data: repeated choices of same consumer over time
- Question: Even if we had data on individual consumer choices, should we estimate models using the individual level data? Or aggregate data from the individual level using the BLP type framework (possible with additional moments)?

Micro and Panel Data

- Individual consumer choices *without* matched consumer characteristics has no new information relative to market level data
- Consumer characteristics not matched to individual choices is just a version of market level data
- We could also have (micro) panel data with repeated observations on the same consumer

Why Micro Data?

- Micro data links individual characteristics to individual consumer choices
  - individuals with children with Volkswagen Golf
  - rich individuals are less sensitive to price
- With market level data, we learn about the marginal distributions of demographics and choices
- With micro data, we learn about their joint distribution

Why Micro Data?

- Using micro data, we can exploit observable variation between demographics and choices to estimate substitution patterns
- Often consumer/choice-specific variables will involve interactions between consumer attributes and product attributes
- For example, many applications have utilized measures of consumer×product-specific distances (consumer distance to retailer/hospital/etc j affects utility from choosing j but not choosing k)
- Within a market, (x_t, p_t, ξ_t) are fixed but choices vary with (observable) consumer attributes
  - Note that this "within market variation" has no endogeneity problem because ξ_t is fixed within a market.
- For how micro data can exactly help in identification, see Berry and Haile (2020)

Why Micro Panel Data?

- See repeated choice for the same consumer over time
- Even the choice sets can change
- Provide additional information about the role of individual characteristics in determining choices (individual fe's ξ_i)
- Note that there can be also state dependence/inertia in choices, might be crucial to take into account
  - Implications on choices, pricing, markups etc...

Estimation Using Micro Data

Consider the following BLP-stype preferences for consumer i, good j in market m at time t:

$$u_{ijmt} = x_{jmt}\beta_{im} + \xi_{jmt} + e_{ijmt} \tag{16}$$

- e_{ijmt} type 1 extreme value distributed
- $\beta_{im} = \beta + \Pi d_{im} + \sigma\nu_{im}$ where d_{im} are demographics (observed at the consumer level in micro data)

Estimation Using Micro Data

Probability that each consumer i choosing j in market m at time t:

$$s_{ijmt} = \int \frac{\exp[\delta_{jmt} + u_{ijmt}(\Pi, \sigma)]}{1 + \sum_k \exp[\delta_{kmt} + u_{ikmt}(\Pi, \sigma)]} dF_\nu(\nu_{im})$$

where

- $\delta_{jmt} = x_{jmt}\beta + \xi_{jmt}$
- $u_{ijmt}(\Pi, \sigma) = x_{jmt}(\Pi d_{im} + \sigma \nu_{im})$

Estimation Using Micro Data

$$s_{ijmt} = \int \frac{\exp[\delta_{jmt} + u_{ijmt}(\Pi, \sigma)]}{1 + \sum_k \exp[\delta_{kmt} + u_{ikmt}(\Pi, \sigma)]} dF_\nu(\nu_{im})$$

- As before, we need to simulate from F_ν
- Unlike before, we observe consumer demographics d_{im} in micro data

Estimation Using Micro Data

$$s_{ijmt} = \int \frac{\exp[\delta_{jmt} + u_{ijmt}(\Pi, \sigma)]}{1 + \sum_k \exp[\delta_{kmt} + u_{ikmt}(\Pi, \sigma)]} dF_\nu(\nu_{im})$$

- The presence of the unobserved quality measure ξ_{jmt} in $\delta_{jmt} = x_{jmt}\beta + \xi_{jmt}$ again implies that MLE cannot be used to estimate the model
- Instead, the contraction mapping introduced in BLP can be used to transform the equation into a linear equation for δ_{jmt} and the coefficients β
- Estimation using GMM/MSM

Micro Data: Key Takeaways

- If you have micro data, it's almost always a good idea to use them!
- With micro data you need to think about whether/to what extent the individual specific data you have is enough to capture the richness of choices
- Need to think about what type of unobservables you might need to make the model rich enough

Estimation Using Micro Data: Some Useful Literature

- Berry, S. and Haile, P. A. 2020. Nonparametric Identification of Differentiated Products Demand Using Micro Data. NBER Working Paper No. 27704.
- Ho, K. 2006. The welfare effects of restricted hospital choice in the US medical care market. J. Appl. Econ., 21: 1039-1079.
- Ho, K. 2009. Insurer-Provider Networks in the Medical Care Market. American Economic Review, 99 (1): 393-430.
- Simulation-based estimation: Train's book "Discrete Choice Methods with Simulation"

Petrin (2002): Combination of Aggregate and Micro Data

Questions

- What are the welfare gains from innovation?
- How much of these gains are captured by consumers and the innovator?
- How big is the extent of first-mover advantage (for the innovator) and the profit cannibalization that took place both initially by the innovator and later by the imitators?

Questions

- What are the welfare gains from innovation?
- How much of these gains are captured by consumers and the innovator?
- How big is the extent of first-mover advantage (for the innovator) and the profit cannibalization that took place both initially by the innovator and later by the imitators?
- The results suggest that the introduction generated large welfare gains for consumers and surplus for the innovator at the expense of the other producers

Setting: Case of the Chrysler's Minivan

Source: Wikipedia

New Alternative for Station Wagons

Source: Wikipedia

Introduction of Minivans

TABLE 3
FAMILY VEHICLE SALES AS A PERCENTAGE OF TOTAL VEHICLE SALES:
U.S. AUTOMOBILE MARKET, 1981–93

Year	Minivans (1)	Station Wagons (2)	Sport-Utilities (3)	Full-Size Vans (4)	Minivans and Station Wagons (5)	U.S. Auto Sales (Millions) (6)
1981	.00	10.51	.58	.82	10.51	7.58
1982	.00	10.27	.79	1.17	10.27	7.05
1983	.00	10.32	3.51	1.04	10.32	8.48
1984	1.58	8.90	5.51	1.20	10.48	10.66
1985	2.32	7.33	6.11	1.05	9.65	11.87
1986	3.63	6.70	5.73	.85	10.43	12.21
1987	4.86	6.47	6.44	.73	11.33	11.21
1988	5.97	5.14	7.18	.69	11.11	11.76
1989	6.45	4.13	7.47	.61	10.58	11.06
1990	7.95	3.59	7.78	.27	11.54	10.51
1991	8.29	3.05	7.80	.29	11.34	9.75
1992	8.77	3.07	9.33	.39	11.84	10.12
1993	9.93	3.02	11.66	.29	12.95	10.71

Micro BLP (Petrin 2002)

Empirical approach

- Random coefficient discrete choice model (BLP)
- Additional micro moments
  - Augment the market-level data on sales and characteristics with information that relates the average demographics of consumers to the characteristics of the products they purchase
  - Information on aggregates of purchasers of new cars

Combining Aggregate and Micro Data

- The extra information plays the same role as consumer-level data, allowing estimated substitution patterns and (thus) welfare to directly reflect demographic-driven differences in tastes for observed characteristics
- For example, observing average family size conditional on the purchase of a minivan and asking the model to reproduce this same average helps to more precisely identify the taste term relating families and minivans
- Similarly, matching probabilities of purchase conditioned on different income levels helps to identify income effects

Combining Aggregate and Micro Data

- Market/product level data (as in BLP):
  - Characteristics and market shares for all new vehicles marketed in the United States from the years 1981 to 1993 (with sales over 1,000 vehicles)
  - Household demographics for representative sample of U.S. population from the Consumer Expenditure Survey (CEX)
- Additional micro data from CEX:
  - Demographics of purchasers of new vehicles linked to the vehicles they purchase for smaller sample (30,000 households)

Utility Specification

The utility specification of consumer i for good j

$$u_{ijt} = \alpha_i \ln(y_i - p_{jt}) + x_{jt}\beta_{it} + \xi_{jt} + e_{ijt} \tag{17}$$

Utility Specification

The utility specification of consumer i for good j

$$u_{ijt} = \alpha_i \ln(y_i - p_{jt}) + x_{jt}\beta_{it} + \xi_{jt} + e_{ijt} \tag{17}$$

- $$\alpha_i = \begin{cases} \alpha_1 & \text{if } y_i < \hat{y}_1 \\ \alpha_2 & \text{if } \hat{y}_1 \leq y_i < \hat{y}_2 \\ \alpha_3 & \text{if } y_i \geq \hat{y}_2 \end{cases}$$

  where $\hat{y}_1$ and $\hat{y}_2$ divide the U.S. population into three equally sized groups ordered by income

Utility Specification

The utility specification of consumer i for good j

$$u_{ijt} = \alpha_i \ln(y_i - p_{jt}) + x_{jt}\beta_{it} + \xi_{jt} + e_{ijt} \tag{17}$$

- $$\alpha_i = \begin{cases} \alpha_1 & \text{if } y_i < \hat{y}_1 \\ \alpha_2 & \text{if } \hat{y}_1 \leq y_i < \hat{y}_2 \\ \alpha_3 & \text{if } y_i \geq \hat{y}_2 \end{cases}$$

  where $\hat{y}_1$ and $\hat{y}_2$ divide the U.S. population into three equally sized groups ordered by income

- $$\beta_{it}^k = \begin{cases} \beta^k + \gamma^k \ln(fs_i)v_{it} & k \text{ for minivan, station wagons (+2 other car type) dummies} \\ \beta^k + \gamma^k v_{it} & \text{for all other } k \end{cases} \tag{18}$$

  For example, γ^k for the minivan dummy is a taste shifter that allows families of different sizes (fs_i) to value minivans differently

Utility Specification

The utility specification of consumer i for good j

$$u_{ijt} = \alpha_i \ln(y_i - p_{jt}) + x_{jt}\beta_{it} + \xi_{jt} + e_{ijt} \tag{19}$$

- $$\alpha_i = \begin{cases} \alpha_1 & \text{if } y_i < \hat{y}_1 \\ \alpha_2 & \text{if } \hat{y}_1 \leq y_i < \hat{y}_2 \\ \alpha_3 & \text{if } y_i \geq \hat{y}_2 \end{cases}$$
- $$\beta_{it}^k = \begin{cases} \beta^k + \gamma^k \ln(fs_i) v_{it} & k \text{ for minivan, station wagons (+2 other car type) dummies} \\ \beta^k + \gamma^k v_{it} & \text{for all other } k \end{cases} \tag{20}$$
- e_{ijt} type 1 extreme value distributed, v_{it} e.g. normal/log-normal distributed

Moments

- In BLP, we had the following moment condition in the GMM objective:

$$E[\xi_{jt}|z_{jt}] = 0 \rightarrow E[\xi'_{jt}z_{jt}] = 0 \tag{21}$$

- Additional micro moments in GMM to match the average model predictions to the observed averages in the data for the outcomes

Additional Micro Moments

- Petrin also considers the following moments:
  1. Probability of purchasing a new vehicle, given income (discretized into three bins)

$$E(\text{purchasing a new vehicle}|y_i < \hat{y}_1)$$

$$E(\text{purchasing a new vehicle}|\hat{y}_1 \leq y_i < \hat{y}_2)$$

$$E(\text{purchasing a new vehicle}|y_i \geq \hat{y}_2)$$

  2. Average family size (fs_i) given the type of vehicle purchased

$$E(fs_i|\text{purchase } j), j = \text{minivan, station wagon, etc}$$

  3. Probability the head of household is 30–60 years old, given the type of vehicle purchased

Additional Micro Moments

- How would you predict e.g. $E(fs_i|\text{purchase } j), j =$minivan, station wagon, etc?

Additional Micro Moments

- How would you predict e.g. $E(fs_i|\text{purchase } j), j =$minivan, station wagon, etc?
- Denote "i purchases j in market t" by $a_{it} = j$
- More specifically, $E(fs_i|a_{it} = j) \equiv E(fs_i|a_{it} = j, \delta_t, \theta_2)$
- Bayes rule:
  $E(fs_i|a_{it} = j, \delta_t, \theta_2) = \int fs_i dF(fs_i|a_{it} = j, \delta_t, \theta_2) = \int fs_i \frac{P(a_{it}=j|fs_i,\delta_t,\theta_2)}{P(a_{it}=j|\delta_t,\theta_2)} dF(fs_i)$
- For estimation, just stack the micro moments (matching model predictions to data) with the original IV moments (21) and run BLP as usual
- Also have a supply side model of oligopolistic competition (pricing decisions, no formal model of entry/exit/innovation)

Demand Estimates

TABLE 4
PARAMETER ESTIMATES FOR THE DEMAND-SIDE EQUATION

Variable	OLS Logit (1)	Instrumental Variable Logit (2)	Random Coefficients (3)	Random Coefficients and Microdata (4)
		A. Price Coefficients (α's)		
α_1	.07	.13	4.92	7.52
	(.01)**	(.01)**	(9.78)	(1.24)**
α_2			11.89	31.13
			(21.41)	(4.07)**
α_3			37.92	34.49
			(18.64)**	(2.56)**

Demand Estimates

TABLE 5
RANDOM COEFFICIENT PARAMETER ESTIMATES

	RANDOM COEFFICIENTS (γ's)	
VARIABLE	Uses No Microdata (1)	Uses CEX Microdata (2)
Constant	1.46	3.23
	(.87)*	(.72)**
Horsepower/weight	.10	4.43
	(14.15)	(1.60)**
Size	.14	.46
	(8.60)	(1.07)
Air conditioning standard	.95	.01
	(.55)*	(.78)
Miles/dollar	.04	2.58
	(1.22)	(.14)**
Front wheel drive	1.61	4.42
	(.78)**	(.79)**
γ_{mi}	.97	.57
	(2.62)	(.10)**
γ_{sw}	3.43	.28
	(5.39)	(.09)**
γ_{su}	.59	.31
	(2.84)	(.09)**
γ_{pv}	4.24	.42
	(32.23)	(.21)**

NOTE.—The OLS and instrumental variable models assume that these random coefficients are zero. Standard errors are in parentheses. A quadratic time trend is included in all specifications. The subscript *mi* stands for minivan, *sw* for station wagon, *su* for sport-utility, and *pv* for full-size passenger van.
* Z-statistic >1.
** Z-statistic >2.

Counterfactuals

- Benefits or harms of the minivan introduction

Counterfactuals

- Benefits or harms of the minivan introduction
  - to consumers
  - to the innovator (Chrysler)
  - to competitors (e.g. General Motors with station wagon)
  - to imitating firms (almost all other)

Counterfactuals

- Measure changes in welfare (consumer welfare, profits) from the minivan introduction

Counterfactuals

- Measure changes in welfare (consumer welfare, profits) from the minivan introduction
  - Take minivan out of the market
  - Simulate new market shares, profits, and welfare
  - For consumers:
    - draw D_i, v_i, e_{ij}
    - calculate choice and utility with full choice set (with minivans) and with reduced choice set (without minivans)

Counterfactuals

- Measure changes in welfare (consumer welfare, profits) from the minivan introduction
  - Take minivan out of the market
  - Simulate new market shares, profits, and welfare
  - For consumers:
    - draw D_i, v_i, e_{ij}
    - calculate choice and utility with full choice set (with minivans) and with reduced choice set (without minivans)
    - calculate compensating variation CV_i as a measure changes in consumer welfare from the introduction of the minivan
    - CV_i=dollar amount the consumer would need to be just indifferent between the equilibrium with minivans and the one without them

Profits

TABLE 11

CHANGE IN INDUSTRY AND BIG THREE TOTAL VARIABLE PROFITS WITH THE ADVENT OF MINIVANS

YEAR	INDUSTRY	CHRYSLER		FORD		GM	
1984	−.21%	\$202.5	14.38%	−\$31.8	−1.16%	−\$155.8	−1.50%
1985	−.13%	\$259.1	13.99%	−\$37.4	−1.29%	−\$171.0	−1.63%
1986	.14%	\$201.1	12.42%	\$54.7	1.84%	−\$119.9	−1.09%
1987	.17%	\$346.1	23.27%	−\$22.8	−.66%	−\$174.5	−2.14%
1988	.65%	\$504.1	32.50%	−\$24.7	−.70%	−\$235.4	−2.90%

NOTE.—Dollar figures are given in millions. The numbers are computed using the model to estimate profits both with minivans in the market and with minivans removed from the market (see Sec. V).

TABLE 12

CHRYSLER'S PROFIT DISSIPATION WITH ENTRY OF FORD AND GM MINIVANS

YEAR	CHANGE IN TOTAL VARIABLE PROFITS	
1985	−\$6.06	−.16%
1986	−\$22.72	−1.99%
1987	−\$42.35	−2.25%
1988	−\$55.68	−2.63%

NOTE.—These profit changes are computed using the model (see the text).

Welfare

TABLE 13

CHANGE IN U.S. WELFARE FROM THE MINIVAN INNOVATION, 1984–88 ($ Millions)

Year	Compensating Variation	Change in Producer Profits	Welfare Change
1984	367.29	−36.68	330.61
1985	625.04	−25.07	599.97
1986	439.93	27.30	467.23
1987	596.59	29.75	626.34
1988	775.70	110.24	885.94
Total	2,804.55	105.54	2,910.09

NOTE.—Computations were done using 1982–84 CPI-adjusted dollars.

Part 1+2: Summary

- BLP give us both a statistical **estimator** and an **algorithm** to obtain estimates.
- Attractive for many differentiated products markets
- Flexible substitution patterns, addressing endogeneity concerns
- Computationally demanding, progress on these challenges
  - Alternative algorithms (MPEC), improvements in the BLP NFP algorithm
  - Current best practice: pyBLP

Part 1+2: Summary

- Many variations of the model in the literature, e.g. using different data structures
  - Aggregate data (original BLP)
  - Micro data
  - Panel data
  - Hybrid
- Often a good idea to use micro data if you have them
- Help in identification, provide additional variation to pin down substitution patterns/heterogeneity in consumer preferences
- Later: identification, more about IVs, supply side

Any Questions?

Email: tanja.saxell@vatt.fi

References

- Berry, S. 1994. Estimating Discrete-Choice Models of Product Differentiation. The RAND Journal of Economics, 25(2), 242-262.
- Berry, S. and Haile, P. A. 2020. Nonparametric Identification of Differentiated Products Demand Using Micro Data. NBER Working Paper No. 27704.
- BLP = Berry, Steven, Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. Econometrica, 63(4), 841–890.
- Conlon, C. and J. Gortmaker. 2020. Best Practices for Differentiated Products Demand Estimation with pyblp. The RAND Journal of Economics, *forthcoming*.
- Dubé, J.-P., Fox, J.T. and Su, C.-L. 2012. Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation. Econometrica, 80: 2231-2267.

References

- Ho, K. 2006. The welfare effects of restricted hospital choice in the US medical care market. J. Appl. Econ., 21: 1039-1079.
- Ho, K. 2009. Insurer-Provider Networks in the Medical Care Market. American Economic Review, 99 (1): 393-430.
- Nevo, A. 2000. Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry. The RAND Journal of Economics, vol. 31, no. 3, pp. 395–421.
- Nevo, A. 2000. "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," Journal of Economics & Management Strategy, 9(4), 513–548.
- Petrin, A. 2002. Quantifying the Benefits of New Products: The Case of the Minivan. Journal of Political Economy, 110(4).
- Waldfogel, J. 2003. Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets. RAND Journal of Economics, 34, 557–568.