- Vertical Model Example: (Shaked and Sutton, Bresnahan 97)

$$u_{ij} = v_i x_j - p_j$$

  - x_j is the quality of the product
  - v_i is the consumer's taste for quality (ie willingness to pay)
- Rather than model utility directly as a function of price, it might be preferable to model it as a function of expenditures on other products and then derive the indirect utility as a function of price
- There are J alternatives in market, indexed by $j = 1, ..., J$
- At each purchase occasion, each consumer divides her income on (at most) one of the alternatives, and on an outside good z:

$$\max_{j,z} U_i(x_j, z) s.t. p_j + p_z z = y_i$$

Random Utility Model (RUM)(McFadden)

- U_{ij}^{*} usually specified as a sum of two parts

$$U_{ij}^{*}(x_j, p_j, p_z, y_i) = V_{ij}(x_j, p_j, p_z, y_i) + \varepsilon_{ij}$$

  - ε_{ij} i.i.d. across products and consumers; represents consumer tastes (observed by consumer but not by the researcher)

- What does it mean for tastes to be represented by product and consumer specific random terms?
  - product chosen is random from the researchers point of view
  - McFadden won the Nobel Prize for this in 2000
  - Assumptions about distribution of the ε_{ij}'s determines choice probabilities

- The probability that consumer i buys product j is

$$D_{ij}(p_1, ...p_j, p_z, y_i) = \Pr ob\left\{\varepsilon_{i0}, ..., \varepsilon_{ij} : U_{ij}^{*} > U_{ik}^{*}, \text{ for } j \neq k\right\}$$

One Product Example

- Buy good 1 (and not outside good $j = 0$) if

$$V_{i0} + \varepsilon_{i0} \leq V_{i1} + \varepsilon_{i1} \iff V_{i0} - V_{i1} \leq \varepsilon_{i1} - \varepsilon_{i0}$$

- Conditional Probit: errors distributed $N(0,1)$
- where probability good 1 is purchased conditional on covariates

$$\Pr(\varepsilon_{i1} - \varepsilon_{i0} \geq V_{i0} - V_{i1} = 1 - \Phi(V_{i0} - V_{i1})$$

- Conditional Logit: errors distributed EV (double exponential)

$$F(\varepsilon) = e^{-e^{-\varepsilon}}$$

- with market share

$$s_{i1} = \frac{\exp(V_{i1})}{\exp(V_{i0}) + \exp(V_{i1})}$$

Independence of Irrelevant Alternatives (IIA)

- ratio of choice prob (odds ratio) does not depend on the number of alternatives available

$$\frac{s_{ij}}{s_{in'}} = \frac{\exp(V_{ij})}{\exp(V_{in'})}$$

- Red bus/blue bus problem: Walk or take red bus
  - If consumer walks half the time then $s_{iW} = s_{iRB} = 0.5$
  - odds ratio walk/RB=1
- Introduce a red bus
  - odds ratio between walk/BB is 1
- But buses are perfect substitutes
  - new choice prob should be $s_{iW} = 0.5; s_{iRB} = s_{iBB} = 0.25$
  - new odds ratio should be walk/RB=2
- IIA is especially troubling if want to predict penetration of new products

Price Elasticities of Demand

- Let $V_{ij} = \alpha p_j + x_j \beta$
- then own and cross-price elasticity of demand between two products

$$\begin{aligned} \frac{\partial s_{ij}}{\partial p_j} &= -\alpha s_{ij}(1 - s_{ij}) \\ \frac{\partial s_{ij}}{\partial p_k} &= \alpha s_{ij} s_{ik} \end{aligned}$$

- Is it concerning that they depend only on the market shares of the products?
- Yes, do not depend on the degree to which products have similar characteristics

Price Elasticities of Demand

- Let $V_{ij} = \alpha p_j + x_j \beta$
- then own and cross-price elasticity of demand between two products

$$\begin{aligned} \frac{\partial s_{ij}}{\partial p_j} &= -\alpha s_{ij}(1 - s_{ij}) \\ \frac{\partial s_{ij}}{\partial p_k} &= \alpha s_{ij} s_{ik} \end{aligned}$$

- Is it concerning that they depend only on the market shares of the products?
- Yes, do not depend on the degree to which products have similar characteristics

Price Elasticities of Demand

- Let $V_{ij} = \alpha p_j + x_j \beta$
- then own and cross-price elasticity of demand between two products

$$
\begin{aligned}
\frac{\partial s_{ij}}{\partial p_j} &= -\alpha s_{ij}(1 - s_{ij}) \\
\frac{\partial s_{ij}}{\partial p_k} &= \alpha s_{ij} s_{ik}
\end{aligned}
$$

- Is it concerning that they depend only on the market shares of the products?
- Yes, do not depend on the degree to which products have similar characteristics

Counter-intuitive substitution patterns:

- Not only from the distributional logit assumption
- Due to assumption that the only variance in consumer tastes comes through the i.i.d. product-specific terms ε_{ij}
- Since i.i.d., there is no source of correlation in consumer tastes across similar products
- Changes to allow for more intuitive substitution patterns
  - Generalized EV models (GEV, Nested logit)
  - Mixtures of logits (K types of logit parameters)
  - Product differentiation model (Bresnahan, Stern, Trajtenberg 1997)
  - Random Coefficients Model of Demand (Berry, Levinsohn, and Pakes)

Berry, Levinsohn, Pakes (BLP) 1995 ECMA

- Method for estimating demand in differentiated product markets using aggregate data (ie only data on market shares not individual choices)
- endogenous prices and random coefficients.
- consistent estimation even with imperfect competition
- To motivate framework consider Berry (RAND, 1994)
  - There are $i = 1, ..., I = \infty$ agents in $t = 1, ..., T$ markets who choose among $j = 1, ..., J$ mutually exclusive alternatives
- K observed product characteristics: $X_{jt} = (x_{j,1,t}, ..., x_{j,K,t})'$
  - Product characteristics/choice sets may evolve over markets
- one unobserved product characteristics: $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$
  - ξ_j is a permanent component for j; ξ_t is a common shock and $\Delta\xi_{jt}$ is a product/time specific shock for j

- Consumer i's indirect utility is given by

$$U_{ijt} = \underbrace{X_{jt}'\beta - \alpha p_{jt} + \xi_{jt}}_{\equiv \delta_{jt}} + \varepsilon_{ijt}$$

- Derive market-level (aggregate) share expression from individual model of discrete-choice
- ε_{ijt} are iid EV so the probability i chooses j is given by

$$s_{ijt} = \frac{\exp(\delta_{jt})}{1 + \sum_k \exp(\delta_{kt})}$$

- Aggregate market shares for product j are (weighted) sum of individual choice probabilities (M is the market size)

$$s_{jt} = \frac{1}{M}[Ms_{ijt}] = \frac{\exp(\delta_{jt})}{1 + \sum_k \exp(\delta_{kt})}$$

- construct the moment condition

$$\frac{1}{J}\sum_{j=1}^{J} E[(\delta_{jt}(S) - X_{jt}\beta + \alpha p_{jt})Z] \equiv Q_{jt}(\alpha, \beta)$$

- Can estimate α and β by minimizing

$$\min_{\alpha,\beta} Q_{jt}(\alpha, \beta)^2$$

- Why does this work?
- As J gets large, by the law of large numbers $Q_{jt}(\alpha, \beta)$ converges to $E[(\delta_{jt} - X_{jt}\beta + \alpha p_{jt})Z]$
- If there exist appropriate instruments Z then at the true values of the parameters this expectation is equal to zero
- Hence, the (α, β) that minimizes $Q_{jt}(\alpha, \beta)^2$ should be close to the true parameter values

- construct the moment condition

$$\frac{1}{J}\sum_{j=1}^{J} E[(\delta_{jt}(S) - X_{jt}\beta + \alpha p_{jt})Z] \equiv Q_{jt}(\alpha, \beta)$$

- Can estimate α and β by minimizing

$$\min_{\alpha,\beta} Q_{jt}(\alpha, \beta)^2$$

- Why does this work?
- As J gets large, by the law of large numbers $Q_{jt}(\alpha, \beta)$ converges to $E[(\delta_{jt} - X_{jt}\beta + \alpha p_{jt})Z]$
- If there exist appropriate instruments Z then at the true values of the parameters this expectation is equal to zero
- Hence, the (α, β) that minimizes $Q_{jt}(\alpha, \beta)^2$ should be close to the true parameter values

What are appropriate instruments?

- IV for j should be correlated with p_j but not with structural error ξ_j
- Usual demand case: cost shifters
  - but we have cross-sectional (across products) data, so we require IV to vary across products within a market
- Example: cars, one natural cost shifter are wages in Michigan
- Here doesn't work because its the same across all products
  - if ran 2SLS with wages in Michigan as IV, first stage regression of price on wage would yield the same predicted price for all products

Random Coefficient Logit

- A well-known solution to problems with logit is to interact product and consumer characteristics (second contribution of BLP)
- ε is EV, like the logit, but β_i, α_i are consumer-specific random coefficients from a parametric distribution

$$u_{ij} = X_j\beta_i - \alpha_i p_j + \xi_j + \varepsilon_{ij}$$

  - Variance is added to the term α or β so substitution patterns can become more reasonable
  - Assume that β_i and α_i are distributed across consumers according to some parametric distribution
  - The own- and cross-derivatives are more flexible. Why?

Intuition of the Estimation Algorithm

- The model is one of individual behavior, yet only aggregate data is observed.
- We can still estimate the parameters that govern the distribution of individuals
  - compute predicted individual behavior and aggregate over individuals, for a given value of the parameters,
  - obtain predicted market shares
- We then choose the values of the parameters that minimize the distance between these predicted shares and the actual observed shares
- The metric under which this distance is minimized is not the straightforward sum of least squares
- rather it is the metric defined by the instrumental variables and the GMM objective function
- It is this last step that somewhat complicates the estimation procedure

Overview GMM Estimation Algorithm

- Guess a parameter vector θ
- Solve for δ and therefore ξ
- Interact ξ and instruments Z these are the moment conditions $Q(\theta)$
- Calculate the objective function $f(\theta) = Q'AQ$ for some

  positive definite A how far is $Q(\theta)$ from zero?
- Guess a new parameter and try to minimize f
- Variance of $\widehat{\theta}$ includes variance in data across products and simulation error as well as any sampling variance in the observed market shares
- Can simplify the algorithm since δ in linear in some parameters – see NEVO (JEMS 2000) for details

Steps for Simulation

- There are essentially four steps (plus an initial step) to follow in computing the estimates:

0 prepare the data including draws from the distribution of v and D

1 for a given value of θ and δ compute the market shares

2 for a given θ, compute the vector δ that equates the market shares computed in Step 1 to the observed shares;

3 for a given θ, compute the structural error term (as a function of the mean valuation computed in Step 2), interact it with the instruments, and compute the value of the objective function;

4 search for the value of θ that minimizes the objective function computed in Step 3.

Supply Side

- Simplest models of product differentiation involve single product firms each producing a differentiated product
- We could begin by specifying a demand system for this set of related products, together with cost functions and an equilibrium notion.
- The usual assumption is Nash-in-prices.
- Profits of firm j are given by

$$\pi_j(p) = p_j q_j(p) - C_j(q_j(p))$$

- The first order condition is

$$q_j + (p_j - mc_j)\frac{\partial q}{\partial p_j} = 0$$

- We can rewrite as $p_j = mc_j + b_j(p)$

- where the price-cost markup is

$$b_j(p) = \frac{q_j}{\left|\frac{\partial q}{\partial p_j}\right|}$$

- Assume that marginal cost is

$$mc_j = w_j\eta + \lambda q_j + \omega_j$$

  - where w_j might consist of X and input prices and q is output
  - ω_j is a supply shock unobserved to the econometrician
- Combining, the FOC is then

$$p_j = w_j\eta + \lambda q_j + b_j(p) + \omega_j$$

- If demand parameters are known then the markup is known and can estimate by IV methods (eg 2SLS) where IV are demand-side variables
- Alternatively mc and demand can be estimated together

Multi-Product Firms

- Non-cooperative oligopolistic Bertrand competition
- Firm f produces a subset $j \in \mathcal{J}_f$ of the products: Profits

$$\sum_{j \in \mathcal{J}_f} (p_j - mc_j) \mathcal{M} s_j(p, X, \xi; \theta)$$

  - where $\mathcal{M}$ is market size
  - s_j is the simulated aggregate market share
- Marginal costs

$$mc_j = w_j' \eta + \omega_j$$

- Any product must have prices that satisfy

$$s_j(p, a) + \sum_{r \in \mathcal{J}_f} (p_r - mc_r) \frac{\partial s_r(p, a)}{\partial p_j} = 0$$

- Given demand can solve for marginal costs and for ω_j

- In vector form, the J FOC are

$$s - \Omega(p - mc) = 0$$

- Notice this implies a markup equation $p - mc = \Omega^{-1}s$
- Ω is called the ownership matrix (of dimension JxJ)
  - Each element takes on the value of $\partial s_r(p, a)/\partial p_j$ for every product that the firm owns
- To estimate the FOC think of estimating the equation

$$mc_j = p_j - b_j(p, x, \xi; \theta) = w_j'\eta + \omega_j$$

- Just as in estimating demand, estimates of the parameters η can be obtained from orthogonality conditions between ω and appropriate instruments

- In vector form, the J FOC are

$$s - \Omega(p - mc) = 0$$

- Notice this implies a markup equation $p - mc = \Omega^{-1}s$
- Ω is called the ownership matrix (of dimension JxJ)
  - Each element takes on the value of $\partial s_r(p, a)/\partial p_j$ for every product that the firm owns
- To estimate the FOC think of estimating the equation

$$mc_j = p_j - b_j(p, x, \xi; \theta) = w_j'\eta + \omega_j$$

- Just as in estimating demand, estimates of the parameters η can be obtained from orthogonality conditions between ω and appropriate instruments

Nevo: Measuring Market Power in the RTE Cereal Industry

- The ready-to-eat (RTE) cereal industry is characterized by high price-to-cost margins (PCM) and high concentrations
- Antitrust authorities accused firms of collusive pricing behavior
- Nevo tests whether this is the case by estimating the price-cost margin (PCM) and decomposing it into 3 sources:
  1. that due to product differentiation
  2. that due to multiproduct form pricing and
  3. that due to price collusion
- Overview of methodology:
  - use the BLP framework to estimate brand-level demand.
  - use demand estimates and different pricing rules to back out PCMs.
  - compare PCMs against crude measures of actual PCM to separate the different sources of the markup

Model and Data

- Indirect utility is

$$u_{ijt} = \alpha_i p_{jt} + X_j \beta_i + \xi_j + \Delta\xi_{jt} + \varepsilon_{ijt}$$

  - uses brand dummy variables (ξ_j) to capture the mean characteristics of RTE cereal
  - once brand dummy variables are included in the regression, the error term is the unobserved city-quarter specific deviation from the overall mean valuation of the brand :structural error is the change in ξ_j over time (denoted $\Delta\xi_{jt}$)
- Cannot use BLP Type Instruments
  - there is no variation in each brand's observed characteristics over time and across cities
  - only variation in IVs from characteristics is due to changes in choice set of available brands
  - proposes alternative IV to separate the exogenous variation in prices (due to differences in mc) and endogenous variation (due to differences in unobserved valuation)

IVs with brand dummies

- Exploit the panel structure of the data (similar to those used by Hausman (1996))
- The identifying assumption is that, controlling for brand specific means and demographics, city-specific valuations are independent across cities (but are allowed to be correlated within a city)
- Given this assumption, the prices of the brands in other cities are valid IV's.
    - prices of brand j in two cities correlated due to the common mc
    - but due to the independence assumption will be uncorrelated with market specific valuation.
    - One could potentially use prices in all other cities and all quarters as instruments
- Independence assumption may not hold (for instance, if there is a national demand shock related to health of cereal)

Identifying Collusive Behavior

- Recall the markup is given by

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

- With single product firms the price of each brand is set by a profit-maximizing firm that considers only the profits from that brand. In this case the ownership matrix will be diagonal
- With multi-product firms, firms set the prices of all their products jointly. In this case some off diagonals will be non-zero
- With collusion, firms act as one firm which owns all products (ie joint profit-maximization of all the brands). In this case the ownership matrix will have no zeros
- Nevo estimates parameters under different definitions of the ownership matrix

Identifying Collusive Behavior

- Recall the markup is given by

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

- With single product firms the price of each brand is set by a profit-maximizing firm that considers only the profits from that brand. In this case the ownership matrix will be diagonal
- With multi-product firms, firms set the prices of all their products jointly. In this case some off diagonals will be non-zero
- With collusion, firms act as one firm which owns all products (ie joint profit-maximization of all the brands). In this case the ownership matrix will have no zeros
- Nevo estimates parameters under different definitions of the ownership matrix