Discrete Choice Methods in Demand Estimation

School of Economics
University of East Anglia

Farasat A.S. Bokhari
f.bokhari@uea.ac.uk
http://www.uea.ac.uk/economics

IO Summer School, 2019

- **Rationale and Aims:** The course is designed to complement other PhD courses in the program, thereby meeting a perceived demand for increased quantitative skills in the area of economic modeling. Such skills are increasingly necessary where researchers are required to provide evidence based policy recommendations based on cutting edge methodologies on complex data sets. The purpose of this course is to prepare students to understand, and to apply, structural econometric methods from industrial organization to their research. Structural econometrics uses economic theory and statistical methods to derive underlying unknown primitives of economic models, which are then often used to guide policy. This module helps advanced students (second year or above) gain hands-on skills when carrying out such research.
- **Learning Outcomes:** By the end of this course, students will be able to estimate demand systems using aggregate sales data on a set of differentiated products via a discrete choice methods including logit, nested logit and random coefficients logit models

- **Topics:** The following is an approximate outline of main topics covered during this course.
  - Overview
    1. Typical problems in estimation
    2. Product vs. characteristics space approach
  - Estimation in Characteristics Space
    1. Random Utility Model and BLP
    2. Logit and Nested Logit
    3. Random Coefficients Logit
    4. Estimation Details

- **Readings:** There is no single text for this course. I will provide lecture notes as we go along but they draw heavily from several sources. Primary ones are listed below.
  - Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.
  - Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.
  - Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.
  - Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy*, 9(4):513–548.

  Other useful material to consult includes Ackerberg et al. (2007), Cameron and Trivedi (2005) (Chapter 6), Train (2003) (Chapters 3 & 9), Hausman et al. (1994), and Reiss and Wolak (2007).

MODULE INFORMATION
ACKNOWLEDGEMENTS

- Demand systems often form the bedrock upon which empirical work in industrial organization rests
- A fundamental issue is to measure market power, which is measured by the price-cost margin

$$L \equiv \frac{p - mc}{p} \qquad \text{(L = Lerner Index)} \tag{1.1}$$

- Lerner Index is a measure of a firm's market power (index ranges from a high of 1 to a low of 0, where for a perfectly competitive firm with $p = mc$, value of Lerner index is zero)
- But cost is often not observed – the "new empirical industrial organization" (NEIO) literature is motivated by this data problem
- General idea – measure the demand side and back out the price cost margins
- How?

- Consider the monopolist's maximization problem,

$$\max_p pq(p) - c(q(p)) \tag{1.2}$$

FOC imply

$$q(p) + p\frac{\partial q(p)}{\partial p} = \frac{\partial c(q(p))}{\partial q}\frac{\partial q(p)}{\partial p} = mc(q(p))\frac{\partial q(p)}{\partial p} \tag{1.3}$$

At the optimal price

$$(p^* - mc(q(p^*))) = -\left.\frac{q(p)}{\partial q(p)/\partial p}\right|_{p=p^*} \tag{1.4}$$

or equivalently,

$$L = \frac{p^* - mc(q(p^*))}{p^*} = -\frac{1}{\eta(p^*)} \tag{1.5}$$

where $\eta(p^*) = \frac{p}{q(p)} \left.\frac{\partial q(p)}{\partial p}\right|_{p=p^*}$ is the price elasticity of demand

Why Demand Estimation?
Single product monopolist

- Inferring costs:

$$L \equiv \frac{p^* - mc(q(p^*))}{p^*} = -\frac{1}{\eta(p^*)}$$

- If the monopolist is pricing optimally, then estimate/knowledge of elasticity η allows us to infer marginal cost mc
- Similarly, if there was a cost shock, and if we have inferred the marginal cost mc, then we figure out its impact on price (assuming the firm still behaves optimally) from the Lerner condition

$$p = mc + \frac{1}{(\partial q(p)/\partial p)} q(p)$$

- Price is equal to marginal cost plus a markup
- The markup depends on the curvature of the demand curve (if demand is perfectly elastic, as in the case of the perfect competition, then p = mc)
- Thus, if we can estimate demand elasticity, we can back out the markups
- Idea extends to oligopoly as well

Why Demand Estimation?

Demand Curves – multiproduct/ologopoly

- When there are differentiated products, we want to estimate the system of demand equations and infer the markups using the full cross-elasticity matrix

$$\begin{aligned} q_1 &= q_1(p_1, p_2, \dots p_j, \dots, p_J, \boldsymbol{X}_1; \xi_1, \boldsymbol{\theta}_1) \\ q_2 &= q_2(p_1, p_2, \dots p_j, \dots, p_J, \boldsymbol{X}_2; \xi_2, \boldsymbol{\theta}_2) \\ &\vdots \\ q_j &= q_j(p_1, p_2, \dots p_j, \dots, p_J, \boldsymbol{X}_3; \xi_j, \boldsymbol{\theta}_j) \end{aligned}$$

  where $j = 1, \dots, j, \dots, J$ represent the J different related products and $\boldsymbol{\theta}_j$ are the paraments in the j-th demand function $q_j(\cdot)$ that need to be estimated

- Elasticity matrix is represented by

$$\boldsymbol{\eta} = \begin{bmatrix} \eta_{11} & \eta_{12} & \dots & \eta_{1j} \\ \eta_{21} & \eta_{22} & \dots & \eta_{2j} \\ & & \vdots & \\ \eta_{J1} & \eta_{J2} & \dots & \eta_{JJ} \end{bmatrix} \qquad \text{where} \qquad \eta_{ji} = \frac{\partial q_j(\cdot)}{\partial p_i} \frac{p_i}{q_j(\cdot)}$$

- Example ...

- Example ...
- Say there are just three related products ... $J = 3$ and demand is specified in log-log form (aka Cobb-Douglas)

$$\begin{aligned} lnq_1 &= \alpha_{10} + \beta_{11} lnp_1 + \beta_{12} lnp_2 + \beta_{13} lnp_3 + \gamma_{14} X_1 + \eta_1 \\ lnq_2 &= \alpha_{20} + \beta_{21} lnp_1 + \beta_{22} lnp_2 + \beta_{23} lnp_3 + \gamma_{24} X_2 + \eta_2 \\ lnq_3 &= \alpha_{30} + \beta_{31} lnp_1 + \beta_{32} lnp_2 + \beta_{33} lnp_3 + \gamma_{34} X_3 + \eta_3 \end{aligned}$$

then the elasticity matrix is constructed from the β parameters

$$\boldsymbol{\eta} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \qquad \text{where} \qquad \eta_{ji} = \frac{\partial q_j(\cdot)}{\partial p_i} \frac{p_i}{q_j(\cdot)} = \frac{\partial lnq_j}{\partial lnp_i} = \beta_{ji}$$

- Note that with just three products, the elasticity matrix in the example above requires estimating at least nine parameters from the demand **system** above

Why Demand Estimation?
Supply Side – multiproduct/ologopoly

- Let the demand for the j^{th} product be given by $q_j = q_j(\mathbf{p})$ (ignoring other variables that enter the j-th demand function and where $\boldsymbol{p}$ is a vector of all related prices)
- If there are L firms, and the lth firm produces a subset $\mathfrak{L}_l$ of the products, then it maximizes its joint profit over those products as

$$\Pi_l = \sum_{r \in \mathfrak{L}_l} (p_r - mc_r) q_r(\mathbf{p})$$

  where mc_r is the (constant) marginal cost of the r^{th} product in the relevant range (and the sum is over the products owned by firm l).
- First order conditions for profit maximization (Nash-Bertand competition) give

$$q_j(\mathbf{p}) + \sum_{r \in \mathfrak{L}_l} (p_r - mc_r) \frac{\partial q_r(\mathbf{p})}{\partial p_j} = 0$$

- This equation gives the equivalent of the lerner conditon we saw earlier for the monopolist ...

- FOC is

$$q_j(\mathbf{p}) + \sum_{r \in \mathfrak{L}_l} (p_r - mc_r) \frac{\partial q_r(\mathbf{p})}{\partial p_j} = 0$$

- In matrix form we can re-write as

$$\mathbf{p} = \mathbf{mc} + \boldsymbol{\Omega}^{-1} \boldsymbol{q}(\boldsymbol{p})$$

- where
    - $\boldsymbol{\Omega}$ is defined such that $\Omega_{jr} = -\Theta_{jr} \frac{\partial q_r(\mathbf{p})}{\partial p_j}$
    - $\boldsymbol{\Theta}$ is 1/0 joint "ownership" matrix with ones in the leading diagonals and in r, i position if these products are produced by the same firm and zeros everywhere else
    - If (some) firms are colluding, they would maximize the sum of profits associated with their joint products – implies changing the terms of the ownership matrix
- Equivalent of the Lerner equation seen before for the multiproduct and multi firm case
- States that price is equal to marginal cost plus a markup
- With the demand system specified and its parameters estimated, we can compute markups and hence back out the marginal costs

Why Demand Estimation?
Examples

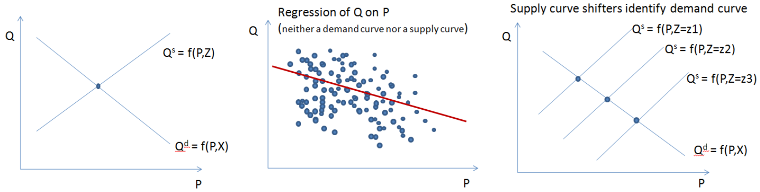
- These demand estimates can be used in a variety of different contexts, including
  - estimating market power
  - backing out marginal costs
  - simulations to predict post-merger prices
  - simulations to predict prices under collusion/cartel
  - estimating the value of new goods (via changes in consumer surplus)
  - to answer other policy questions such as impact of allowing direct to consumer advertising or parallel trade, or accessing the impact of tariffs/import duties etc.
- The process thus begins with estimating a system of demand equations
- Example
  - Estimate an appropriate demand system
  - Compute the markups and the implied marginal costs using the ownership matrix
  - Recompute (simulate) prices based on the demand parameters, marginal cost and a new joint ownership structure (merger or collusion), i.e. by changing the terms of the $\boldsymbol{\Theta}$ ownership matrix

- Demand models often suffer from the endogeneity problem
  - Endogeneity means when in an econometric equation, a right hand side is correlated with the error term
  - In demand models, this is because the prices on the right hand side are typically correlated with the error term
  - A consequence of that is that it violates one of the classical assumptions of the OLS regression theory and hence leads to biased estimates of the demand parameters
- The Problem – Consider an equation such as

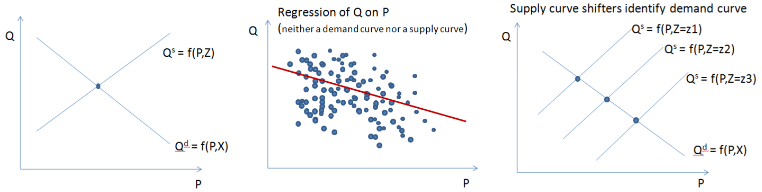
$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

where the interest is in knowing the value of β_2.

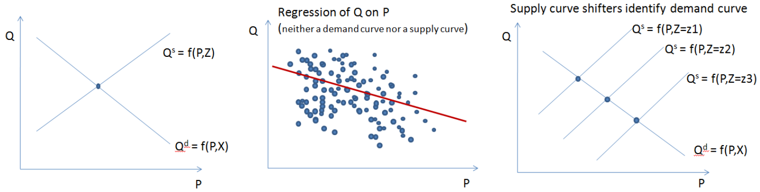
  - If ($E[X_{2i}, u_i] \neq 0$) then simple regression based methods will produce biased estimates such that $E(\widehat{\beta_2}) \neq \beta_2$.
  - This is because $E[X_{2i}, u_i] \neq 0$ (crucial assumption in OLS) due to
    - measurement error of X_2
    - omitted variable(s) X_3 correlated with both Y and X_2
    - simultaneity – i.e., where X_2 and Y are jointly determined



- In typical demand analysis with n products
  - Quantity demanded is function of own price, price of related products and other demand shifters, $Q_i^d = f(p_1, p_2, \ldots, p_i, \ldots, p_n, X_i)$.
  - The supply curve Q_i^s is also a function of own price and marginal cost $Q_i^s = f(p_i, C_i)$.
  - The *observed* price and quantity (or shares) are jointly determined via market clearing (demand equals supply, $Q_i^d = Q_i^s$).
  - Regression of quantity on prices (even after holding other variables constant) will result in neither the estimates of demand curve nor of the supply curve.
  - Demand curve can be *identified* via variables that shift the supply curve (e.g. cost of production).



- The Cure – For each endogenous variable such as X_2, find a variable (instrument) Z such that
  - it is relevant (i.e., $E[X_{2i}, Z_i] \neq 0$)
  - it is valid (i.e., $E[Y_i, Z_i] = 0$)
- The IV procedure – In two easy steps
  - Regress X_{2i} on Z_i and obtain predicted values of X_2 (say $\widehat{X}_2$)
  - Regress Y on $\widehat{X}_2$ – coefficient on X_2 is now an unbiased estimate of β_2



- Instruments
  - To estimate demand curves, we need at least n relevant and valid instruments $(Z_1, Z_2, \ldots, Z_n)$.
  - C_i enter the supply function and hence are relevant (i.e., $E[p_i, C_i] \neq 0$).
  - C_i do not enter the demand function and hence are valid (i.e., $E[Q_i, C_i] = 0$).
  - Good News: Can use the (marginal) costs C_i of the products as instruments for the prices.
  - Bad News: Data on marginal costs by product line is often not available.
- Need some different type of instruments to estimate demand curves.

- Prices are often endogenous (due to simultaneity) ...
- Consider a very simple linear demand/supply model for a single homogenous product over T markets, where aggregate demand/supply relations are given by

$$\begin{aligned} q_t^d &= \beta_{10} + \gamma_{12} p_t + \beta_{11} x_{1t} + \xi_{1t}, \\ p_t &= \beta_{20} + \gamma_{22} q_t^s + \beta_{22} x_{2t} + \xi_{2t}, \\ q_t^s &= q_t^d \end{aligned}$$

- error terms are such that[1]

$$\begin{aligned} E(\xi_{1t}|\mathbf{x}_t) = 0, E(\xi_{2t}|\mathbf{x}_t) &= 0, \\ E(\xi_{1t}^2|\mathbf{x}_t) = \sigma_1^2, E(\xi_{2t}^2|\mathbf{x}_t) &= \sigma_2^2 \\ E(\xi_{1t}\mathbf{x}_t) = 0, E(\xi_{2t}\mathbf{x}_t) &= 0, \\ \text{and } E(\xi_{1t}\xi_{2t}|\mathbf{x}_t) &= 0 \end{aligned}$$

where $\mathbf{x}_t = [1 \; x_{1t} \; x_{2t}]$

[1] Since we have already made the stronger assumption that $E(\xi_{1t}|\mathbf{x}_t) = 0$, technically we do not need to explicitly make the assumption that $E(\xi_{1t}\mathbf{x}_t) = 0$, since the latter is implied by the former assumption of zero conditional mean due to law of iterated expectations. Nonetheless, I include it just to be clear.

- Prices are often endogenous (due to simultaneity) ...
- solve for the reduced form equilibrium values of q^* and p^* – dropping subscript t, we get

$$q^* = \frac{\beta_{10} + \beta_{20}\gamma_{12}}{1 - \gamma_{12}\gamma_{22}} + \frac{\beta_{11}}{1 - \gamma_{12}\gamma_{22}} x_1 + \frac{\gamma_{12}\beta_{22}}{1 - \gamma_{12}\gamma_{22}} x_2 + \frac{\xi_1 + \gamma_{12}\xi_2}{1 - \gamma_{12}\gamma_{22}}$$
$$p^* = \frac{\beta_{20} + \beta_{10}\gamma_{22}}{1 - \gamma_{12}\gamma_{22}} + \frac{\beta_{11}\gamma_{22}}{1 - \gamma_{12}\gamma_{22}} x_1 + \frac{\beta_{22}}{1 - \gamma_{12}\gamma_{22}} x_2 + \frac{\gamma_{22}\xi_1 + \xi_2}{1 - \gamma_{12}\gamma_{22}}$$

- p^* is a function of ξ_1 (and ξ_2) and hence an OLS estimation of the demand equation above (regress q on p, x_1) will result in an inconsistent estimate of γ_{12} and other parameters

- Prices are often endogenous (due to simultaneity) ...
- Useful to explicitly compute the conditional covariance between p and ξ_1
- Note that conditional on $\mathbf{x}_t$,

$$p^* - E(p^*) = \frac{\gamma_{22}\xi_1 + \xi_2}{1 - \gamma_{12}\gamma_{22}}$$

$$\text{and } \xi_1 - E(\xi_1) = \xi_1$$

Thus

$$\text{cov}(p, \xi_1) = \frac{\gamma_{22}}{1 - \gamma_{12}\gamma_{22}}\sigma_1^2 + \frac{\mathrm{E}(\xi_1\xi_2)}{1 - \gamma_{12}\gamma_{22}}$$

- Even if the error terms across the two equations were uncorrelated ($\mathrm{E}(\xi_{1t}\xi_{2t}|\mathbf{x}_t = 0$), the covariance between p and ξ_1 would still not be zero
- On the other hand, if γ_{22} is zero, q does not appear in the supply equation, i.e., it is a triangular system of equations and OLS estimation is fine as long as $\mathrm{E}(\xi_{1t}\xi_{2t}|\mathbf{x}_t) = 0$
- For completeness – complete system of equations, i.e., the number of equations are equal to the number of endogenous variables – we also require that $\gamma_{12} \neq 1/\gamma_{22}$.

Demand Models
Product vs Characteristics Space

- Product Space
  - Consumers have preferences over products
  - Usual utility maximization problem
  - Leads to demand at the product level
  - In that sense, demand analysis in product space is more natural (or at least more familiar)
- Characteristics Space
  - Views products as bundles of characteristics
  - Consumers have preferences over those characteristics
  - Each individual's demand for a given product is just a function of the characteristics of the product
- We can think of a set of products (Toyota Minivan, Lexus SUV, etc.) or we can think of them as a collection of various properties (horsepower, size, color, etc.)
- In general, demand systems in characteristic space are approximations to product space demand systems and hence, we can either model consumers as having preferences over products, or over characteristics (note that not all of the characteristics need to be observed and may form part of the error term)

- Considerations
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - New Goods
  - Cross elasticities

- Considerations
  - Dimensionality of Products
    - For large number of products (say $J = 50$), the product space approach leads to the dimensionality problem and may require grouping/nesting these products.
    - Consider a system of demand equations

$$\mathbf{q} = D(\mathbf{p}, \mathbf{z}; \boldsymbol{\theta}, \boldsymbol{\xi}) \tag{1.6}$$

      where $\mathbf{q}$ is a $J \times 1$ vector of quantities, $\mathbf{p}$ is a vector of prices, $\mathbf{z}$ is a vector of exogenous variables that shift demand, $\boldsymbol{\theta}$ are the parameters to be estimated, and $\boldsymbol{\xi}$ are the error terms
    - In a system with J products, even with some simple and restrictive forms, the number of parameters to estimate is large
    - If $D(\cdot)$ is linear so that $D(\mathbf{p}) = \mathbf{A}\mathbf{p}$ where $\mathbf{A}$ is a $J \times J$ matrix of slope coefficients, then there are J^2 parameters to estimate (plus additional ones due to the exogenous variables $\mathbf{z}$). Can impose restrictions but may not solve the problem [(1) Slutsky equation: $\frac{\partial q_j}{\partial p_i} = \frac{\partial h_j}{\partial p_i} - q_i \frac{\partial q_j}{\partial y}$; (2) Engle aggregation: $\sum_j s_j \eta_{jy} = 1$; (3) Cournot aggregation: $\sum_j s_j \eta_{ji} = -s_i$ where η_{ji}]
  - Dimensionality of Characteristics
  - New Goods
  - Cross elasticities

- Considerations
  - Dimensionality of Products
  - Dimensionality of Characteristics
    - By contrast, if we can reduce J products to just a few K characteristics, and the preferences over those characteristics are, say normally distributed, then we have to estimate K means and $K(K+1)/2$ covariances. *If* there were no unobserved characteristics, then $K(1+(K+1)/2)$ parameters would suffice to analyze own and cross price elasticities for all J goods.
    - If there are too many characteristics (K is large), then the problem of too many parameters re-appears as in the product space, and we need data on each of these characteristics. A solution is to model some of them as unobserved characteristics – but this leads to the endogeneity problem if the unobserved characteristics (think product quality) are correlated with the price, which they usually are.
  - New Goods
  - Cross elasticities

- Considerations
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - New Goods
    - If we are interested in the counterfactual exercise to assess the welfare impact of a new introduction in an ex-ante period (say a new proposed generic drug or a me-too drug), it is difficult to do so in the product space (we can do it using ex-post data though), but it is easier to do this exercise using the characteristic space approach. This is because if we have estimated the demand system using the characteristic approach, and we know the proposed characteristics of the new good, we can, in principle, analyze what the demand for the new good would be. Note however that if the new good is totally different from products already in the market, i.e., has very different (and new) properties, characteristics space approach may not help either (e.g., could we have predicted the demand for laptops based on the characteristics of desktop computers, or for a new drug which proposes treatment of a formally un-treatable disease?)
  - Cross elasticities

- Considerations
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - New Goods
  - Cross elasticities
    - Most of the characteristics space estimation, at least on aggregate data, does not easily lend to analyzing products which are used in bundles or as complements. This is an on-going area of research.

- Depending on the context and the question, a researcher needs to be careful about choosing the appropriate estimation methodology
- Earlier empirical work focused on specifying representative consumer demand systems such that they allowed for various substitution patterns, and were consistent with economic theory
  - Linear Expenditure model (Stone, 1954)
  - the Rotterdam model (Theil, 1965; and Barten 1966)
  - or the more flexible ones such as the Translog model (Christensen, Jorgenson, and Lau, 1975) and the Almost Ideal Demand System (Deaton and Muellbauer, 1980a)
- We will focus on the discrete choice models – logit, nested logit, random coefficients logit – based on works by Berry (1994) and Berry et al. (1995) (henceforth BLP)

- Consumer chooses a single product from a finite set of goods
- Each product is defined as a bundle of attributes (including price, which is a special attribute), and consumers have preferences over these attributes
- Consumers can have different relative preferences, which gives rise to the random coefficients models, and they choose the product that maximizes their utility subject to the usual constraints
- This leads to different choices by different consumers
- Aggregate demand is then derived as the sum over individuals and depends on the entire distribution of consumer preferences

- Indirect utility for individual n for product j in market t is given by

$$u_{njt} = U(\mathbf{x}_{jt}, \xi_{jt}, y_{nt} - p_{jt}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{njt}; \boldsymbol{\theta}_n), \qquad \text{for } j = 0, 1, 2, \ldots, J \tag{2.1}$$

  - ‘outside good’ is numbered 0 (when the consumer does not purchase any of the observed products)
  - price of the outside good is often considered to be exogenous
  - vector $\mathbf{x}_{jt}$ and random term ξ_{jt} are the observed and unobserved (to the econometrician, but not to the consumer) product characteristics and do not vary over consumers
  - product characteristics, multiplied by the parameters $\boldsymbol{\theta_n}$ determine the level of utility for consumer n

- Indirect utility for individual n for product j in market t is given by

$$u_{njt} = U(\mathbf{x}_{jt}, \xi_{jt}, y_{nt} - p_{jt}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{njt}; \boldsymbol{\theta}_n), \qquad \text{for } j = 0, 1, 2, \ldots, J \tag{2.1}$$

  - vectors $\boldsymbol{d}_{nt}$ and $\boldsymbol{\nu}_{nt}$ are vectors of observed and unobserved sources of differences in consumer tastes
  - they do not enter the utility function directly, but rather enter into the model by changing the value of the parameters of interest for each consumer
  - $\boldsymbol{d}_{nt}$ may be a vector of observed demographics (income, family size etc.), that effect the parameters (marginal valuations) of product characteristics by individual, and change the value of $\boldsymbol{\theta}$ for each attribute of the product by individual n
  - for each product attribute (including price) there is an additional randomness to the marginal valuation by individuals and is captured by $\boldsymbol{\nu}_{nt}$
  - accounts for other unobserved person specific characteristics that affect their marginal valuation for an observed product characteristic – e.g., number of dogs a family owns affects their marginal valuation of the size of a car

- Indirect utility for individual n for product j in market t is given by

$$u_{njt} = U(\mathbf{x}_{jt}, \xi_{jt}, y_{nt} - p_{jt}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{njt}; \boldsymbol{\theta}_n), \qquad \text{for } j = 0, 1, 2, \ldots, J \tag{2.1}$$

  - if $\mathbf{x}_{jt}$ is a $k-1$ vector of observed characteristics, then $\boldsymbol{\nu}_{nt}$ is a vector of length k
  - the coefficients $\boldsymbol{\theta}_n$ depend on $\boldsymbol{d}_{nt}$ and $\boldsymbol{\nu}_{nt}$
  - ϵ_{njt} is a mean-zero stochastic term that enters directly into the utility of product j for consumer n
  - for each consumer, $\boldsymbol{\epsilon}_{nt} = (\epsilon_{n0t}, \epsilon_{n1t}, \ldots, \epsilon_{nJt})$ is a vector of error terms with the length equal to the number of products
  - y_{nt} is the consumers income, but is often subsumed into either $\boldsymbol{\nu}$ or in $\boldsymbol{d}$, so that utility is modeled explicitly depending on prices, i.e., $u_{njt} = U(\mathbf{x}_{jt}, \xi_{jt}, p_{jt}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{njt}; \boldsymbol{\theta}_n)$
  - utility of the outside good is denoted as $u_{n0t} = U(\mathbf{x}_{0t}, \xi_{0t}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{n0t}; \boldsymbol{\theta})$ and is normalized to zero

- Consumer n will choose product j when $u_{njt} \geq u_{nlt}$ for all $l = 0, 1, \ldots, J$ and $l \neq j$
- Differences in consumer choices arise only due to differences in the marginal valuations $\boldsymbol{\theta}_n$ (which are themselves functions of $\boldsymbol{d}_{nt}$ and $\boldsymbol{\nu}_{nt}$), and the idiosyncratic terms ϵ_{njt}, a consumer can be described as a tuple $(\boldsymbol{d}, \boldsymbol{\nu}, \boldsymbol{\epsilon})$
- The set $\mathbb{A}_{jt}$ defines characteristics of the individuals that choose brand j in market t

$$\mathbb{A}_{jt}(\mathbf{x}_t, \mathbf{p}_t; \boldsymbol{\theta}) = \{(\boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{n0t}, \epsilon_{n1t}, \ldots, \epsilon_{nJt}) \,|\, u_{njt} > u_{nlt} \qquad \forall\, l = 0, 1, 2 \ldots J,\ l \neq \tag{2.2}$$

where $\mathbf{p}_t = (p_{0t}, \ldots, p_{Jt})'$ and $\mathbf{x}_t = (\mathbf{x}_{0t}, \ldots, \mathbf{x}_{Jt})'$

- Market share of product j is just the probability weighted sum of individuals in the set $\mathbb{A}_{jt}$
- Let $F(\boldsymbol{d}, \boldsymbol{\nu}, \boldsymbol{\epsilon})$ be the population joint distribution function, then the market share of product j is the integral of this distribution over the mass of individuals in the region $\mathbb{A}_{jt}$,

$$s_{jt}(\mathbf{x}, \mathbf{p}; \boldsymbol{\theta}) = \int_{\mathbb{A}_{jt}} dF(\boldsymbol{d}, \boldsymbol{\nu}, \boldsymbol{\epsilon}). \tag{2.3}$$

If the size of the market is M (total number of consumers) then the aggregate demand for the jth product is $Ms_{jt}(\mathbf{x}, \mathbf{p}; \boldsymbol{\theta})$

- Let the indirect utility for consumer n for product j in market t be given by

$$\begin{aligned} &u_{njt} = \alpha_n(y_n - p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta}_n + \xi_{jt} + \epsilon_{njt}, \text{ where} \\ &n = 1, \ldots, N, \qquad j = 0, 1, \ldots, J, \qquad t = 1, 2, \ldots, T, \text{ and} \\ &\boldsymbol{\beta}_n = \boldsymbol{\beta}, \qquad \alpha_n = \alpha, \qquad \text{for all } N \end{aligned} \tag{3.1}$$

- where
  - $\mathbf{x}_{jt}$ is a $k-1$ dimensional vector of observable characteristics (which may be varying by markets)
  - ξ_{jt} is a *scalar* that summarizes the unobservable (to the econometrician) product characteristics
  - neither of these terms varies over consumers
  - also, no variation in tastes across consumers and the terms $\boldsymbol{d}_{nt}$ and $\boldsymbol{\nu}_{nt}$ do not enter this model (but later on will make $\boldsymbol{\beta}_n$ and α_n functions of $\boldsymbol{d}_n$ and $\boldsymbol{\nu}_n$ mentioned earlier)
  - outside option (product 0) is normalized by assuming that the price and other characteristics are zero for this option so that

$$u_{n0t} = \alpha y_n + \epsilon_{n0t} \tag{3.2}$$

- Utility function in (3.1) can be written more compactly as just

$$u_{njt} = \alpha y_n + \delta_{jt} + \epsilon_{njt}, \tag{3.3}$$

where $\delta_{jt} \equiv \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$ is the **mean utility** for product j in market t

- Since income is common to all options, and consumers only differ in the terms ϵ, the set of individuals choosing product j is given by

$$\mathbb{A}_{jt}(\boldsymbol{\delta}_t(\mathbf{x}_t, \mathbf{p}_t; \alpha, \boldsymbol{\beta})) = \{(\epsilon_{n0t}, \epsilon_{n1t}, \dots \epsilon_{nJt})|u_{njt} > u_{nlt} \qquad \forall\, l = 0, 1, 2 \dots J,\ l \neq j\} \tag{3.4}$$

where $\boldsymbol{\delta}_t = (\delta_{0t}, \dots, \delta_{Jt})'$, and $\mathbf{p}_t$ and $\mathbf{x}_t$ are defined as before

- Assume ϵ_{njt} are independently and identically distributed (iid) and follow a Type-1 extreme value distribution, given by

$$f(\epsilon) = \exp(-\epsilon)\exp(-\exp(-\epsilon)) \text{ and } \qquad F(\epsilon) = \exp(-\exp(-\epsilon)), \tag{3.5}$$

where $f(\epsilon)$ and $F(\epsilon)$ are the PDF and CDF of the random variable ϵ

- If ϵ_{njt} are iid Type-1 extreme value distribution, then market share of product j (and the probability that individual n chooses product j) is

$$s_{jt}(\boldsymbol{\delta}_t) = \int_{\mathbb{A}_{jt}} dF(\boldsymbol{\epsilon}) = \frac{\exp(\delta_{jt})}{\sum_{j=0}^{J} \exp(\delta_{jt})}. \tag{3.6}$$

- Since $\delta_{0t} = 0$ (so that $(\exp(\delta_{0t}) = \exp(0) = 1)$, the share equation becomes

$$\begin{aligned} s_{jt} &= \frac{\exp(\delta_{jt})}{1 + \sum_{j=1}^{J} \exp(\delta_{jt})} \\ s_{0t} &= \frac{1}{1 + \sum_{j=1}^{J} \exp(\delta_{jt})} = 1 - \sum_{j=1}^{J} s_{jt}. \end{aligned} \tag{3.7}$$

- Since $s_{jt}/s_{0t} = \exp(\delta_{jt})$, and hence

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \tag{3.8}$$

can be estimated using linear regression methods

Standard Logit/Homegenous Tastes

Utility function and market shares

- Since $s_{jt}/s_{0t} = \exp(\delta_{jt})$, and hence

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \tag{3.8}$$

  can be estimated using linear regression methods
- Instead of estimating J^2 number of parameters, we only have to estimate a handful
- Own and cross price elasticities depend on only one parameter α
- The closed (logit) form for the shares is due to both, the extreme value distribution, and the iid assumption
- The independence part of iid, causes serious limitations on the substitution patterns

- The logit model suffers from the property known as the Independence of Irrelevant Alternatives (IIA)
- The (logit) probability that individual n chooses product j is given by (see (3.6))

$$\Pr(j) = \frac{\exp(\delta_j)}{\sum_{j=0}^{J} \exp(\delta_j)} \tag{3.9}$$

The relative probabilities of options j and k are thus

$$\frac{\Pr(j)}{\Pr(k)} = \frac{\exp(\delta_j)}{\exp(\delta_k)} = \exp(\delta_j - \delta_k) \tag{3.10}$$

- Ratio does not depend on characteristics of any other alternative other than those of j and k
- Thus the relative odds of choosing j over k are the same no matter what other alternatives are available or what are the attributes of other alternatives (the values of $\delta' s$)

- IIA leads to substitution patterns that may be unrealistic
- Blue Bus/Red Bus Example
  - A traveler can commute to work either by car (c) or by blue bus (bb)
  - Suppose further that it turns out (for simplicity) that $Pr(bb) = Pr(c) = .5$
  - Say a new type of bus is introduced that is identical in all other respects to the existing blue bus (fare, route, smell, time it takes to get to work, etc.,) except that it is red in color (rb)
  - We expect the new probabilities of travel model would be $Pr(bb) = Pr(rb) = .25$ and $Pr(c) = .5$
  - logit model would predict that the substitution from the two old modes of travel (blue bus or car) to the new mode of travel (red bus) are such that they would depend on the ratio of old probabilities
  - Since the old probabilities were equal, new probabilities for each of the new modes would be $Pr(bb) = Pr(rb) = Pr(c) = 1/3$

- IIA has implications for own and cross elasticities estimated via logit specification for the aggregate demand
- Price elasticities from the model are

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k, \\ \alpha p_{kt} s_{kt} & \text{otherwise} \end{cases} \tag{3.11}$$

- Cross elasticity
  - cross price elasticity between product j and k depends *only* on the prices and shares of product k
  - let Coca Cola = product j; Pepsi Cola = product k; and Orange Cola = product l
  - if the price of Pepsi Cola increases by 1%, then ceteris paribus, the market shares of Coca Cola and Orange Cola will increase by the same proportion regardless of the fact that Coca Colas and Pepsi Cola are more like each other (blue bus/red bus) compared to Orange Cola (car)

- IIA has implications for own and cross elasticities estimated via logit specification for the aggregate demand
- Price elasticities from the model are

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k, \\ \alpha p_{kt} s_{kt} & \text{otherwise} \end{cases} \tag{3.11}$$

- Own elasticity
    - often market shares (when there are many differentiated products) are small
    - own elasticity will be roughly proportional to the price of the product ($\eta_{jjt} \approx -\alpha p_{jt}$ because $(1 - s_{jt}) \approx 1$)
    - if price increases, sensitivity to prices also increases – but people who buy more expensive products may infact be less price sensitive compared to those who buy less expensive products
    - if as the price increases, so does elasticity, it implies that the markups for cheaper priced products will be larger than those with higher priced products (price-costs margin inversely related to own elasticities) – markups are higher for cheaper priced generics compared to the blockbuster patented?

Standard Logit/Homegenous Tastes
Estimation Details

- Despite the earlier noted shortcomings, logit may be ok in some situations – even if not, its easy to estimate and can be a starting point for more elaborate models
- If we have aggregate sales data (quantities and prices), along with product characteristics, equation (3.8) can be estimated by defining the dependent variable y_{jt} as $y_{jt} = \ln(s_{jt}) - \ln(s_{0t})$
- To start, need to estimate the share of the outside good – done by first defining the (potential) size of the market
- Examples
  - Bresnahan et al (1997) define it as the total number of office-based employees
  - BLP define it as total number of households
  - Nevo (2001) defines the potential size of the market as one bowl of cereal per day per person
  - In the example of ADHD drugs considered earlier, one could define it as a 12-hr day-long coverage of a standard dose of ADHD drug – 3×30mg strength of Ritalin IR (a 30mg pill covers about 4hrs of a day) which can be multiplied by a base line candidate population, say 10% of all school aged children (current ADHD prevalence rates of whom only 69% are given any ADHD drugs), and a smaller proportion of the older population

- Thus, first define the potential size of the market M_t
- Next, based on the observed values of $q_{1t}, \ldots, q_{Jt}$, define the shares of the 'inside' goods $s_{1t}, \ldots s_{Jt}$ relative to the market size as

$$s_{jt} = q_{jt}/M_t \qquad j = 1, \ldots, J \text{ for all } t = 1, \ldots, T. \tag{3.12}$$

- Then, share of the outside good per market is just

$$s_{0t} = 1 - \sum_{j=1}^{J} s_{jt} \qquad \forall t \tag{3.13}$$

- With these definitions in place, can estimate equation (3.8) (reproduced below)

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} \equiv \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}, \tag{3.8}$$

via linear regression methods — infact can estimate the equation with data from just one market

- With these definitions in place, can estimate equation (3.8) (reproduced below)

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} \equiv \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}, \tag{3.8}$$

via linear regression methods — infact can estimate the equation with data from just one market

  - let $\mathbf{y}'_t = (y_{1t}, y_{2t}, \ldots, y_{Jt})$ be a row vector (for market t) given by $\mathbf{y}'_t = ([\ln s_{1t} - \ln s_{0t}], [\ln s_{2t} - \ln s_{0t}], \ldots, [\ln s_{Jt} - \ln s_{0t}])$ so that $\mathbf{y}_t$ is a column vector of length J
  - let $\mathbf{p}'_t = (p_{1t}, \ldots, p_{Jt})$ and $\boldsymbol{\xi}'_t = (\xi_{1t}, \ldots, \xi_{Jt})$ be row vectors with J entries for the t-th market
  - since $\mathbf{x}_{jt}$ is a row vector of observable characteristics of product j in market t, i.e., $\mathbf{x}_{jt} = (x_{1jt}, x_{2jt}, \ldots, x_{Kjt})$, thus let $\mathbf{X}'_t = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t}, \ldots, \mathbf{x}'_{jt}, \ldots, \mathbf{x}'_{Jt})$ so that $\mathbf{X}_t$ is a $J \times K$ matrix, such that each row is itself a k dimensional vector of observable product characteristics

Then (3.8) can be written in 'long' form and even estimated with observations from one market t

$$\mathbf{y}_t = (\ln \mathbf{s}_{jt} - \ln s_{0t}) \qquad = \alpha(-\mathbf{p}_t) + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\xi}_t \equiv \boldsymbol{\delta}_t$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}_t = \begin{bmatrix} \ln s_1 - \ln s_0 \\ \ln s_2 - \ln s_0 \\ \vdots \\ \ln s_J - \ln s_0 \end{bmatrix}_t = \alpha \begin{bmatrix} -p_1 \\ -p_2 \\ \vdots \\ -p_J \end{bmatrix}_t + \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1K} \\ x_{21} & x_{22} & \ldots & x_{2K} \\ \vdots & & & \\ x_{J1} & x_{J2} & \ldots & x_{JK} \end{bmatrix}_t \begin{bmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_J \end{bmatrix} \tag{3.14}$$

Standard Logit/Homegenous Tastes

Estimation Details

- Data from multiple markets can be vertically 'stacked'

$$\mathbf{y} = \alpha(-\mathbf{p}) + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi} \equiv \boldsymbol{\delta}$$

$$\begin{bmatrix} \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{J1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Jt} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} y_{1T} \\ y_{2T} \\ \vdots \\ y_{JT} \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} \ln s_{11} - \ln s_{01} \\ \ln s_{21} - \ln s_{01} \\ \vdots \\ \ln s_{J1} - \ln s_{01} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \ln s_{1t} - \ln s_{0t} \\ \ln s_{2t} - \ln s_{0t} \\ \vdots \\ \ln s_{Jt} - \ln s_{01} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \ln s_{1T} - \ln s_{0T} \\ \ln s_{2T} - \ln s_{0T} \\ \vdots \\ \ln s_{JT} - \ln s_{0T} \end{pmatrix} \end{bmatrix} = \alpha \begin{bmatrix} \begin{pmatrix} -p_{11} \\ -p_{21} \\ \vdots \\ -p_{J1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} -p_{1t} \\ -p_{2t} \\ \vdots \\ -p_{Jt} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} -p_{1T} \\ -p_{2T} \\ \vdots \\ -p_{JT} \end{pmatrix} \end{bmatrix} + \begin{bmatrix} \begin{pmatrix} x_{111} & x_{121} & \dots & x_{1K1} \\ x_{211} & x_{221} & \dots & x_{2K1} \\ \vdots & & & \\ x_{J11} & x_{J21} & \dots & x_{JK1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} x_{11t} & x_{12t} & \dots & x_{1Kt} \\ x_{21t} & x_{22t} & \dots & x_{2Kt} \\ \vdots & & & \\ x_{J1t} & x_{J2t} & \dots & x_{JKt} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} x_{11T} & x_{12T} & \dots & x_{1KT} \\ x_{21T} & x_{22T} & \dots & x_{2KT} \\ \vdots & & & \\ x_{J1T} & x_{J2T} & \dots & x_{JKT} \end{pmatrix} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} + \tag{3.15}$$

- Very likely that $\text{cov}(p_{jt}, \xi_{jt}) \neq 0$
  - Endogeneity concern arises in a variety of differentiated products pricing models as well
  - Let the demand for the i^{th} product be given by $q_i = D_i(\mathbf{p}, \mathbf{z}_i; \xi_i)$, where ξ_i is the error term and consists of unobserved product characteristics, and $\mathbf{z}_i$ is the vector of exogenous demand shifters (say the observed product characteristics)
  - If there are L firms, and the lth firm produces a subset $\mathfrak{L}_l$ of the products, then it maximizes its joint profit over these products as

$$\Pi_l = \sum_{r \in \mathfrak{L}_l} (p_r - c_r) q_r(\mathbf{p}, \mathbf{z}_r, \xi_r), \tag{3.16}$$

where c_r is the constant marginal cost of the r^{th} product

- Nash-Bertrand price competition, price p_i of any product i produced by firm l satisfies the first order condition

$$q_i(\mathbf{p}, \mathbf{z}_i; \xi_i) + \sum_{r \in \mathfrak{L}_l} (p_r - c_r) \frac{\partial q_r(\mathbf{p}, \mathbf{z}_r; \xi_r)}{\partial p_i} = 0 \tag{3.17}$$

- The equilibrium price for product i would be a function of its marginal cost and a markup term, and in matrix form (for all equilibrium prices) is given by

$$\mathbf{p} = \mathbf{c} + \boldsymbol{\Omega}^{-1} \boldsymbol{q}(\boldsymbol{p}, \boldsymbol{z}; \boldsymbol{\xi}), \tag{3.18}$$

- where
  - $\boldsymbol{\Omega}$ is defined such that $\Omega_{ri} = -O_{ri} \frac{\partial q_r(\mathbf{p}, \mathbf{z}_r; \xi_r)}{\partial p_i}$
  - $\boldsymbol{O}$ is 1/0 joint ownership matrix with ones in the leading diagonals and in r, i position if these products are produced by the same firm and zeros everywhere else
- The markup term is a function of the same error terms, and hence generally prices will be endogenous, so that OLS/SUR estimation will lead to biased estimates of the demand parameters

- The usual starting place for demand side instruments is to use cost shifters (terms that affect $\mathbf{c}$, such as cost of raw materials) that are uncorrelated with demand shocks
- These can work well for homogenous products, but in the case of differentiated products, we would need costs shifters that vary by individual brands, which are often very difficult to obtain
- Two types of instruments which have grown in popularity (use with caution as may or may not be valid in your application)
  - Berry (1994)/Berry, Levinsohn, and Pakes (BLP) (1995)
  - Hausman et al. (1994)

- Berry (1994) builds on Bresnahan's (1981) assumption that the location of products in a characteristics space is determined prior to the revelation of consumer's valuation of the unobserved product characteristics
- BLP use this assumption to generate a set of instrumental variables: they use the observed product characteristics (excluding price and any other endogenous characteristics of the product), the sums of the values of the same characteristics of other products offered by that firm, and the sums of the values of the same characteristics of products offered by other firms
  - Consider the case when there are two firms, X and Y and each is producing three products A,B,C and D,E,F respectively
  - Suppose further that each of these products have two observable characters, S (say, package size, which is the number of pills in a box) and T (number of times a pill must be taken during a day for a standard diagnosis
  - Then for the price of A, which is produced by firm X, there are 6 potential instruments:
    - S_{AX} and T_{AX} – the values of S and T of product A
    - $S_{BX} + S_{CX}$ and $T_{BX} + T_{CX}$ – the sum of S and T over the firms two other products B and C
    - $S_{DY} + S_{EY} + S_{FY}$ and $T_{DY} + T_{EY} + T_{FY}$ – the sum of S and T over the competitors products D,E and F
  - Similar instruments can be constructed for prices of other products

- Main advantage of this approach (if valid) is that it gives instruments that vary by brands
- Problems arise if the assumption that the observed characteristics are uncorrelated with observed characteristics is not valid
  - for instance, if the observed characteristics are changing over time, and the change in observed characteristics is for the same unobserved factors that determine price
- Another potential issue arises if brand dummies are included in the estimation, since then it must be the case that there is variation in products offered in different markets, else there will be no variation between the instruments in these markets

- A second set of instruments is due to Hausman et al. (1994) and has been used in several papers
- Hausman uses the panel nature of data and the assumption that prices in different areas (cities) are correlated via **common cost shocks**, to use prices from other areas as instruments for prices in a given city and there are **no common demand side shocks** across the two cities
- The identifying assumption is that after controlling for brand specific intercepts and demographics, the city specific valuations of a product are independent across cities but may be correlated within a city over time
- Given this assumption, the prices of the brand in other cities are valid instruments so that prices of brand j in two cities will be correlated due to the common marginal cost, but due to the independence assumption will be uncorrelated with the market specific valuation of the product
- Instruments may be in invalid if the error terms are related due to common demand side shocks across the two cities
  - Example: a national campaign will increase the unobserved valuation of product i in both cities, thus violating the independence assumption

- Very likely that $\text{cov}(p_{jt}, \xi_{jt}) \neq 0$
- Regardless of the instruments used, a first approach to consistent estimation would be to estimate a fixed effects model with dummies for products (and markets)
  - Requires that data be available from multiple markets
  - Thus, with data available from multiple markets, one can estimate via OLS

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_j + \xi_t + \Delta\xi_{jt} \tag{3.19}$$

  where ξ_j is the brand fixed effect and ξ_t is the market fixed effect
  - Identifying assumption for OLS estimation is

$$\mathrm{E}(\Delta\xi_{jt}p_{jt}|\mathbf{x}_{jt}) = 0 \tag{3.20}$$

- Thus, with data available from multiple markets, one can estimate via OLS

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_j + \xi_t + \Delta\xi_{jt} \tag{3.19}$$

- A brand specific dummy variable captures all the observed characteristics of the product that do not vary across markets, as well as the product specific mean of the unobserved characteristics, i.e., $\mathbf{x}_j\boldsymbol{\beta}$, where, note the missing market subscript of t from the vector $\mathbf{x}$
- Thus, the correlation between prices and brand specific mean of unobserved quality is fully accounted for and does not require an instrument
- Once brand specific dummy variables are included in the regression, the error term now is just the market specific deviation from the mean of the unobserved characteristics, and may still require the use of instruments if the condition in equation (3.20) is not true

- Thus, with data available from multiple markets, one can estimate via OLS

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_j + \xi_t + \Delta\xi_{jt} \tag{3.19}$$

- Similarly, if the mean unobserved quality – where the mean is now across all brands – is different by markets, then it too is fully accounted for by the market dummies
- If the subscript t for the markets is in the context of time periods, then this could be because the unobserved quality for all products is improving over time (think computer quality over time)
- If the subscript t is in the cross-sectional setting, then this may or may not make much sense, since by adding such dummies to the equation, the researcher is effectively arguing that the unobserved quality components of all brands in, Hooker, OK, are higher than those in Boring, OR
  - This maybe true if the products under study require some additional local input for providing the product (radio channels with local DJs and ads), or if shipping from long distance affects the quality of all products (fresh food), but not if they are centrally produced (RTE cereals) and shipping does not impact quality

- Two objections to the use of brand dummies
- Use of brand dummies increases the number of parameters to be estimated by J (rather than by J^2) – may not be too serious an issue if the number of markets is large
- A potentially more serious difficulty is that the coefficients $\boldsymbol{\beta}$ cannot be identified if observed characteristics do not vary by markets

Standard Logit/Homegenous Tastes

Instruments and Dummy Variables

- Two objections to the use of brand dummies
- Use of brand dummies increases the number of parameters to be estimated by J (rather than by J^2) – may not be too serious an issue if the number of markets is large
- A potentially more serious difficulty is that the coefficients $\boldsymbol{\beta}$ cannot be identified if observed characteristics do not vary by markets
  - Nevo (2001) points out that infact they can be recovered using minimum distance procedure by regressing the estimated brand dummy variables on the observed characteristics
  - Let $\mathbf{b}_t$ be the $J \times 1$ vector of brand dummies and let $\mathbf{X}_t$ be the $J \times K$ matrix of observed product characteristics and $\boldsymbol{\xi}_t$ be the $J \times 1$ vector of unobserved product qualities, neither of which vary by markets
  - Let also $\hat{\boldsymbol{b}}$ be the estimated values of coefficients ($J \times 1$) of the brand dummies and $\hat{\boldsymbol{V}}_{\boldsymbol{b}}^{\mathbf{-1}}$ their estimated $J \times J$ variance covariance matrix, both of which are available from initially estimating equation (3.19)

- Two objections to the use of brand dummies
- Use of brand dummies increases the number of parameters to be estimated by J (rather than by J^2) – may not be too serious an issue if the number of markets is large
- A potentially more serious difficulty is that the coefficients $\boldsymbol{\beta}$ cannot be identified if observed characteristics do not vary by markets
  - Nevo (2001) points out that infact they can be recovered using minimum distance procedure by regressing the estimated brand dummy variables on the observed characteristics
  - Let $\mathbf{b}_t$ be the $J \times 1$ vector of brand dummies and let $\mathbf{X}_t$ be the $J \times K$ matrix of observed product characteristics and $\boldsymbol{\xi}_t$ be the $J \times 1$ vector of unobserved product qualities, neither of which vary by markets
  - Let also $\hat{\boldsymbol{b}}$ be the estimated values of coefficients ($J \times 1$) of the brand dummies and $\hat{\boldsymbol{V}}_{\boldsymbol{b}}^{-1}$ their estimated $J \times J$ variance covariance matrix, both of which are available from initially estimating equation (3.19)
  - Then, the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ in equation

$$\mathbf{b}_t = \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\xi}_t, \tag{3.21}$$

can be recovered via GLS estimator

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}_t' \hat{\mathbf{V}}_{\mathbf{b}}^{-1} \mathbf{X}_t)^{-1} \mathbf{X}_t' \hat{\mathbf{V}}_{\mathbf{b}}^{-1} \hat{\mathbf{b}}_t, \text{ and } \boldsymbol{\xi}_t = \hat{\mathbf{b}}_t - \mathbf{X}_t \widehat{\boldsymbol{\beta}} \tag{3.22}$$

where the latter is just the calculated value of the residual term from the regression above

- The IIA problem in logit arose from the iid structure of the error terms
- Particulary, while consumers have different rankings of the products, these differences arise only due to the iid shocks to the error term ϵ_{njt}
- One solution to this problem is to make the random shocks to the utility correlated across products by generating correlations through the error term
- An example is the nested logit model in which products are grouped and ϵ_{njt} is decomposed into an iid shock plus a group specific component which results in correlation between products in the same group
- Basic idea is to relax the IIA by grouping products (similar to the grouping idea in multilevel budgeting/AIDS we saw earlier), but within each group we have a standard logit model, and products in different groups have less in common and are not good substitutes

- Let the utility for consumer n for product j in group g be

$$u_{njt} = \delta_{jt} + \zeta_{ngt}(\sigma) + (1-\sigma)\epsilon_{njt}, \tag{3.23}$$

- where
  - $\delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$ is the mean utility for product j common to all consumers (as before)
  - ϵ_{njt} is (still) the person specific iid random shock with extreme value distribution
  - but ζ_{ngt} is the person specific shock that is common to all products in group g
  - The distribution of the group specific random variable ζ_{ngt} depends on the parameter σ so that $\zeta_{ngt}(\sigma) + (1-\sigma)\epsilon_{njt}$ is extreme value
  - If σ approaches zero, the model is reduced to that of the simple logit case discussed earlier while if it approached one, only the nests matter
- Gives a closed form which can be estimated using linear estimation methods

$$\ln(s_{jt}) - \ln(s_{0t}) = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \sigma\ln(s_{jt}/s_{gt}) + \xi_{jt} \tag{3.24}$$

- The additional term $\ln(s_{jt}/s_{gt})$ is the share of product j in group g
- All previous issues (define outside good, use of dummies, instruments etc.) apply here as well
- One difference from the previous case is that even if prices are exogenous, the term $\ln(s_{jt}/s_{gt})$ is endogenous and we need some instrumental variable for it

- Suppose we want to estimate a simple linear model

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + u_t \tag{3.25}$$

- where
  - $\mathbf{x}_t$ is a $1 \times K$ vector (including the constant or the intercept term), $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters and u_t is the usual error term
  - conditional on regressors, error term is mean zero so that $\mathrm{E}[u_t|\mathbf{x}_t] = 0$
- Then we get K equations of the form

$$\mathrm{E}[\mathbf{x}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})] = \mathbf{0} \tag{3.26}$$

- The method of moments (MM) estimator is the solution to the corresponding sample moment conditions

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \mathbf{0} \tag{3.27}$$

which gives the MM estimator of $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}}_{MM} = (\sum_t \mathbf{x}_t'\mathbf{x}_t)^{-1}\sum_t \mathbf{x}_t' y_t = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.28}$$

which is just the OLS estimator

- Say we have an additional set of exogenous variables $\mathbf{z}_t$ that are correlated with $\mathbf{x}_t$ but not with the error terms so that $\mathrm{E}[u_t|\mathbf{z}_t] = 0$
- Then, $\mathrm{E}[(y_t - \mathbf{x}_t\boldsymbol{\beta})|\mathbf{z}_t] = 0$, and as before, we can multiply $\mathbf{z}_t$ with the residual terms to get K unconditional population moment conditions

$$\mathrm{E}[\mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})] = \mathbf{0} \tag{3.29}$$

- Then the MM estimator solves the sample moment conditions given by

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \mathbf{0} \tag{3.30}$$

- If $\dim(\mathbf{z}) = K$, then this yields the MM estimator which is just the IV estimator

$$\widehat{\boldsymbol{\beta}}_{MM} = (\sum_t \mathbf{z}_t'\mathbf{x}_t)^{-1}\sum_t \mathbf{z}_t' y_t = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} \tag{3.31}$$

- If however, $\dim(\mathbf{z}) > K$, (more potential instruments than the original number of regressors) then there is no unique solution – more moment conditions than the number or parameters to be estimated
- We can use the GMM estimator which chooses $\widehat{\boldsymbol{\beta}}$ so as to make the vector $T^{-1}\sum_{t=1}^{T}\mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})$ as small as possible using quadratic loss
- Thus find $\widehat{\boldsymbol{\beta}}_{\text{GMM}}$ which minimizes the function

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}\sum_t \mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})\right]' \boldsymbol{\Phi} \left[\frac{1}{T}\sum_t \mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})\right] \tag{3.32}$$

  where $\boldsymbol{\Phi}$ is a $\dim(\mathbf{z}) \times \dim(\mathbf{z})$ weighting matrix
- In matrix notation define $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ (where $\mathbf{y}$ and $\mathbf{u}$ are $T \times 1$, $\mathbf{X}$ is $T \times K$ and $\boldsymbol{\beta}$ is $K \times 1$ as before), and let $\mathbf{Z}$ be $T \times R$ matrix, then $\sum_{t=1}^{T}\mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \mathbf{Z}'\mathbf{u}$ and (3.32) becomes

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{Z}\right] \boldsymbol{\Phi} \left[\frac{1}{T}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \tag{3.33}$$

  where $\boldsymbol{\Phi}$ is a $R \times R$ full rank symmetric weighting matrix

- First order conditions, $\partial Q(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0}$ for the linear IV case are

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = -2\left[\frac{1}{T}\mathbf{X}'\mathbf{Z}\right]\boldsymbol{\Phi}\left[\frac{1}{T}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right] = \mathbf{0} \tag{3.34}$$

- Then the GMM linear IV estimator and its variance are

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\text{GMM}} &= \left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{y} \\ \mathrm{V}(\widehat{\boldsymbol{\beta}})_{\text{GMM}} &= T\left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\widehat{\mathbf{S}}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X}\right)^{-1} \end{aligned} \tag{3.35}$$

where $\widehat{\mathbf{S}}$ is a consistent estimate of

$$\mathbf{S} = \text{plim}\frac{1}{T}\sum_i\sum_j\left[\mathbf{z}_i' u_i u_j \mathbf{z}_j\right] \tag{3.36}$$

- Different choices of the weighting matrix $\mathbf{\Phi}$ lead to different estimators
- If the model is just identified ($R = K$) and the matrix $\mathbf{X}'\mathbf{Z}$ is invertible, then the choice of the weighting matrix $\mathbf{\Phi}$ does not matter as the GMM estimator is just the IV estimator:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\text{GMM}} &= \left(\mathbf{X}'\mathbf{Z}\mathbf{\Phi}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\mathbf{\Phi}\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{\Phi}^{-1}(\mathbf{X}'\mathbf{Z})^{-1}(\mathbf{X}'\mathbf{Z})\mathbf{\Phi}\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} = \widehat{\boldsymbol{\beta}}_{\text{IV}} \end{aligned} \tag{3.37}$$

- If $R > K$, and the errors are homoscedastic, then $\mathbf{\Phi} = (T^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ and $\widehat{\mathbf{S}}^{-1} = \left[s^2 T^{-1}\mathbf{Z}'\mathbf{Z}\right]$ leads to the usual 2SLS estimator

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\text{GMM}} &= \left(\mathbf{X}'\mathbf{P_z}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{P_z}\mathbf{y}\right) = \widehat{\boldsymbol{\beta}}_{\text{2SLS}} \\ \mathrm{V}(\widehat{\boldsymbol{\beta}}_{\text{GMM}}) &= s^2\left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1} \\ \text{where } \mathbf{P_z} &= \mathbf{Z}(\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z}' \text{ and } s^2 = (T-K)^{-1}\sum_t \hat{u}_t^2 \end{aligned} \tag{3.38}$$

- Alternatively, if errors are heteroscedastic, then instead we can use

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\text{GMM}}) = T\left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\widehat{\mathbf{S}}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right.$$

$$\text{and } \widehat{\mathbf{S}} = T^{-1}\sum_t \hat{u}_t^2 \mathbf{z}_t \mathbf{z}'_t. \tag{3.39}$$

- The **optimal** weighting matrix (optimal in the sense of efficiecy/smallest variance) is one which is proportional to the inverse of $\mathbf{S}$
- The optimal GMM two-step estimator (for the linear IV case) is when $\mathbf{\Phi} = \widehat{\mathbf{S}}^{-1}$

$$\widehat{\boldsymbol{\beta}}_{\text{OGMM}} = \left(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y} \tag{3.40}$$

  - Step 1: Use 2SLS as the first-step to estimate $\widehat{\boldsymbol{\beta}}$ and then compute residuals as in the heteroscedastic case above
  - Step 2: Construct the $\widehat{\mathbf{S}}^{-1}$ and then use it in (3.40) to compute the estimator
- Variance is given by

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\text{OGMM}}) = T\left(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1} \tag{3.41}$$

- This approach extends easily to the general case with other moment conditions
- Let $\boldsymbol{\theta}$ be a $q \times 1$ vector of parameters and $\mathbf{h}(\mathbf{w}, \boldsymbol{\theta})$ be an $r \times 1$ vector function such that at the true value of the parameter $\boldsymbol{\theta}_0$, there are r moment conditions ($r > q$) give by

$$\mathrm{E}\left[\mathbf{h}(\mathbf{w}_t, \boldsymbol{\theta}_0)\right] = \mathbf{0} \tag{3.42}$$

  - where the expectations are not zero if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$
  - the vector $\mathbf{w}_t$ includes all observable variables, including $\mathbf{y}_t$, $\mathbf{x}_t$ and, $\mathbf{z}_t$
- Then the GMM objective function (equivalent of (3.32)) is

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}\sum_t \mathbf{h}(\mathbf{w}_t, \boldsymbol{\theta})\right]' \boldsymbol{\Phi} \left[\frac{1}{T}\sum_t \mathbf{h}(\mathbf{w}_t, \boldsymbol{\theta})\right] \tag{3.43}$$

  and the corresponding first order conditions are

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left[\frac{1}{T}\sum_t^T \frac{\partial \mathbf{h}_t(\widehat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}}\right] \boldsymbol{\Phi} \left[\frac{1}{T}\sum_t^T \mathbf{h}_t(\widehat{\boldsymbol{\theta}})\right] = \mathbf{0} \tag{3.44}$$

  where $\mathbf{h}_t(\boldsymbol{\theta}) = \mathbf{h}(\mathbf{w}_t \boldsymbol{\theta})$

- Note that If $\mathbf{h}_t(\boldsymbol{\theta}) = \mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \mathbf{z}_t' u_t$ then $\partial \mathbf{h}/\partial \boldsymbol{\beta}' = -\mathbf{z}_t'\mathbf{x}_t$ and the earlier results of linear IV follows

- GMM also extends to non-linear models, where the error term u_t may or may not be additively separable
- For instance, $u_t = y_t - g(\mathbf{x}_t; \boldsymbol{\theta})$ where $g(\cdot)$ is some nonlinear function but the error term is additively separable, or non-separable so that $u_t = g(y_t, \mathbf{x}_t; \boldsymbol{\theta})$
- If $\mathrm{E}(u_t|\mathbf{x}_t) \neq 0$ but we have instruments available so that $\mathrm{E}(u_t|\mathbf{z}_t) = 0$, then the moment conditions are $\mathrm{E}(\mathbf{z}_t' u_t) = \mathbf{0}$
- The GMM estimator minimizes the objective function

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}\mathbf{u}'\mathbf{Z}\right] \boldsymbol{\Phi} \left[\frac{1}{T}\mathbf{Z}'\mathbf{u}\right] \tag{3.45}$$

- Unlike the linear case, the first order conditions do not give closed forms for the estimators

- Earlier saw that standard logit can be estimated as a linear equation when the dependent variable is defined as $y_{jt} \equiv \ln s_{jt} - \ln s_{0t}$ and the equation is given as $y_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$
- When the price is correlated with the unobserved heterogeneity term ξ_{jt}, so that $\mathrm{E}(p, \xi) \neq 0$ and we have a set of instruments such that $\mathrm{E}(Z\xi) = 0$, then we can use the GMM/IV methods described in the earlier section to estimate the parameters of the equation
- The linear equation arose out of Berry's (1994) inversion trick
- Useful to work through this again for extending the method to random coefficients model

- Let the observed shares be given by $\mathbf{s}$ so that $\mathbf{s}_t = (s_{0t}, s_{1t}, \ldots, s_{Jt})$ where, as before, $s_{0t} = 1 - \sum_{j=1}^{J} s_{jt}$
- Let also $\boldsymbol{\theta}_1 \equiv \begin{bmatrix} \alpha & \boldsymbol{\beta}' \end{bmatrix}'$) and let model predicted market shares in equation (3.7) be given by $\tilde{\mathbf{s}}$ so that $\tilde{\mathbf{s}}_t = (\tilde{s}_{0t}, \tilde{s}_{1t}, \ldots, \tilde{s}_{Jt})$
- Given a value of $\boldsymbol{\theta}_1$, can compute the model predicted shares as

$$\tilde{s}_{jt} = \frac{\exp(\delta_{jt})}{1 + \sum_{j=1}^{J} \exp(\delta_{jt})} \tag{3.7}$$

- Thus, may want to use NLS methods to find $\boldsymbol{\theta}_1$ so as to minimize the distance between predicted and observed market shares

$$\min_{\boldsymbol{\theta}_1} \sum_{j=1}^{J} \left[s_{jt} - \tilde{s}_{jt}(\alpha, \boldsymbol{\beta}, \xi_{1t}, \xi_{2t}, \ldots, \xi_{Jt}) \right]^2 \tag{3.46}$$

- The econometric error terms $\boldsymbol{\xi}_t$ – unobserved product qualities – enter the predicted market share and are not additively separable. Hence, non-linear least squares methods will not give consistent estimates *even if* prices were not endogenous

University of East Anglia

- Assume that we have a set of M instruments given by matrix $\mathbf{Z}$ with dimensions $JT \times M$ (the jt^{th} row is given by $\mathbf{z}_{jt} = (z_{jt}^{(1)}, z_{jt}^{(2)}, \ldots, z_{jt}^{(M)})$) which are uncorrelated with error terms in the utility model ξ_{jt}
- Then the M moment conditions are given by $\mathrm{E}(\mathbf{z}'_{jt}\xi_{jt}) = \mathbf{0}$
- The key insight comes from the fact that the error terms enter the mean utility linearly ($\delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$), and that they only enter the mean utility and hence one can separate out the ξ_{jt} terms to compute the moment conditions above

$$\frac{1}{J}\sum_{j} z_{jt}^{(m)}\xi_{jt} = \frac{1}{J}\sum_{j} z_{jt}^{(m)}(\delta_{jt} - \mathbf{x}_{jt}\boldsymbol{\beta} + \alpha p_{jt}) \tag{3.47}$$

- Thus want to estimate the parameters $\alpha, \boldsymbol{\beta}$ that minimize the sample moment conditions (or rather their weighted sum of squares)
- But since we cannot observe δ_{jt} we cannot proceed as is
- Berry (1994) suggests a two step approach: first obtain an estimate of δ_{jt}, – call it $\widehat{\delta}_{jt}$ – and insert it into the moment conditions above, and second search for values of $\alpha, \boldsymbol{\beta}$ that minimize the weighted sum of squares of these moment conditions

(1) Figure out the values of δ_{jt}

(A) If we normalize $\delta_{0t} = 0$ and equate the observed shares to model predicted shares, then we have J non-linear equations per market – see logit share equation (3.7) – in J unknowns

$$\begin{aligned} s_{1t} &= \tilde{s}_{1t}(\delta_{1t}, \ldots, \delta_{Jt}) \\ s_{2t} &= \tilde{s}_{2t}(\delta_{1t}, \ldots, \delta_{Jt}) \\ &\vdots \\ s_{Jt} &= \tilde{s}_{jt}(\delta_{1t}, \ldots, \delta_{Jt}) \end{aligned} \tag{3.48}$$

(B) If we can invert this system, we can solve for $\delta_{1t}, \delta_{2t}, \ldots, \delta_{jt}$ as a function of observed shares $s_{1t}, s_{2t}, \ldots, s_{jt}$.

(C) Thus, we now have $\widehat{\delta}_{jt} \equiv \tilde{s}_{jt}^{-1}(s_{1t}, s_{2t}, \ldots, s_{Jt})$, J numbers per market which we can use to carry out step 2 (in the simple logit case, $\widehat{\delta}_{jt} = \ln(s_{jt}) - \ln(s_{0t})$)

(2) With the estimated values of δ_{jt}, use GMM to estimate parameters (in this case, α and $\boldsymbol{\beta}$) so as to minimize (3.47).

(A) Recall that δ_j is the mean utility of product j defined linearly as $\delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$ for all j,

$$\begin{aligned} \delta_{1t} &= \alpha(-p_{1t}) + \mathbf{x}_{1t}\boldsymbol{\beta} + \xi_{1t} \\ \delta_{2t} &= \alpha(-p_{2t}) + \mathbf{x}_{2t}\boldsymbol{\beta} + \xi_{2t} \\ &\vdots \\ \delta_{Jt} &= \alpha(-p_{Jt}) + \mathbf{x}_{Jt}\boldsymbol{\beta} + \xi_{Jt} \end{aligned} \tag{3.49}$$

(B) We can now use the estimated values of $\widehat{\delta}_j$ to calculate the sample moments

$$\frac{1}{J}\sum_j z_{jt}^{(m)}\xi_{jt} = \frac{1}{J}\sum_j z_{jt}^{(m)}(\widehat{\delta}_{jt} - \mathbf{x}_{jt}\boldsymbol{\beta} + \alpha p_{jt}) \tag{3.50}$$

minimize these to calculate the values of α, $\boldsymbol{\beta}$

- In step (1a) above, we equated observed market shares to model predicted market shares
  - In the case of logits, the model predicted market shares take the closed form (3.7) given by $\tilde{s}_{jt} = \exp(\delta_{jt}) / \left[1 + \sum_{j=1}^{J} \exp(\delta_{jt})\right]$
  - In other cases, there will be no closed form available to compute the model predicted market shares and we will need to resort to numerical simulation methods to estimate the model predicted shares
  - Infact, these may be functions of additional parameters (call them $\boldsymbol{\theta}_2$) – thus, equations (3.48) will be of the form

$$s_{jt} = \tilde{s}_{jt}(\delta_{1t}, \ldots, \delta_{Jt}, \boldsymbol{\theta}_2) \tag{3.51}$$

- In steps (1b/1c), we 'inverted' these equations to solve for $\widehat{\delta}_{jt}$
  - In the case of logit, analytical solution was available since $\delta_{jt} = \ln s_{jt} - \ln s_{0t}$
  - More generally, these equations are nonlinear and need to be solved numerically
  - Berry/BLP suggest a contraction mapping (and prove that it converges) for $\boldsymbol{\delta}_t$ given by

$$\boldsymbol{\delta}_t^{h+1} = \boldsymbol{\delta}_t^h + \left[\ln(\mathbf{s}_t) - \ln(\tilde{\mathbf{s}}_t(\boldsymbol{\delta}_t^h; \boldsymbol{\theta}_2))\right] \tag{3.52}$$

  where $\mathbf{s}_t(\cdot)$ is the observed market share, $\tilde{\mathbf{s}}_t(\cdot)$ is the model predicted market share at mean utility $\boldsymbol{\delta}_t^h$ at iteration h and $||\boldsymbol{\delta}_t^{h+1} - \boldsymbol{\delta}_t^h||$ is below some tolerance level

- To sum up, Berry's (1994) two step GMM approach with a matrix of instruments $\mathbf{Z}$ is as follows:
  (1) Compute $\widehat{\delta}_{jt}$
    - Without loss of generality, subsume p_{jt} within $\mathbf{x}_{jt}$ as just another column (a special attribute of product J), and rather than introduce new (unnecessary) notation, redefine $\mathbf{x}_{jt} = \begin{bmatrix} -p_{jt} & \mathbf{x}_{jt} \end{bmatrix}$ – similarly, redefine matrix $\mathbf{X}$ to be inclusive of the price vector so that $\mathbf{X} = \begin{bmatrix} \mathbf{p} & \mathbf{X} \end{bmatrix}$. Also, let $\mathbf{s}_t$ be the vector of observed shares and $\boldsymbol{\theta}_1 = \begin{bmatrix} \alpha & \boldsymbol{\beta}' \end{bmatrix}'$
    - Conveniently, $\widehat{\delta}_{jt} = \ln(s_{jt}) - \ln(s_{0t})$ (in the case of simple logit) and $\widehat{\boldsymbol{\delta}} = \ln(\mathbf{s}) - \ln(\mathbf{s}_0)$
    - Then $\xi_{jt}(\boldsymbol{\theta}_1) = \widehat{\delta}_{jt}(\mathbf{s}_t) - \mathbf{x}_{jt}\boldsymbol{\theta}_1$ – and in matrix notation, $\boldsymbol{\xi}(\boldsymbol{\theta}_1) = \widehat{\boldsymbol{\delta}} - \mathbf{X}\boldsymbol{\theta}_1$

  (2) Define the moment conditions as $\mathrm{E}(\mathbf{Z}'\boldsymbol{\xi}(\boldsymbol{\theta}_1)) = \mathbf{0}$
    - Next, $\min\limits_{\boldsymbol{\theta}_1} \boldsymbol{\xi}(\boldsymbol{\theta}_1)'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\boldsymbol{\xi}(\boldsymbol{\theta}_1)$ where $\boldsymbol{\Phi} = (\mathrm{E}[\mathbf{Z}'\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{Z}])^{-1}$
    - In the case of logit, we have an analytical solution – see equation (3.40) in the GMM section, and replace $\mathbf{y}$ in that equation with $\widehat{\boldsymbol{\delta}}$: $\widehat{\boldsymbol{\theta}}_1 = (\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\widehat{\boldsymbol{\delta}}$
    - Since we don't know $\boldsymbol{\Phi}$, we start with $\boldsymbol{\Phi} = \mathbf{I}$ or $\boldsymbol{\Phi} = (\mathbf{Z}'\mathbf{Z})^{-1}$, get an initial estimate of $\boldsymbol{\theta}_1$, use this to get residuals, and then recompute $\boldsymbol{\Phi} = (\mathrm{E}[\mathbf{Z}'\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{Z}])^{-1}$ to get the new estimates of $\boldsymbol{\theta}_1$
- We will use this 2 step approach explicitly in the next model

Random Coefficients Logit
Heterogenous Tastes

- Let the utility be given by

$$u_{njt} = \alpha_n(y_n - p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta}_n + \xi_{jt} + \epsilon_{njt}, \text{ where}$$
$$n = 1, \dots, N, \qquad j = 0 \dots, J, \qquad t = 1 \dots, T \tag{4.1}$$

- where

$$\begin{bmatrix}\alpha_n \\ \boldsymbol{\beta}_n\end{bmatrix} = \underbrace{\begin{bmatrix}\alpha \\ \boldsymbol{\beta}\end{bmatrix}}_{\boldsymbol{\theta}_1} + \underbrace{\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n}_{\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}}$$
$$= \begin{bmatrix}\alpha \\ \boldsymbol{\beta}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Pi}_\alpha \\ \boldsymbol{\Pi}_{\boldsymbol{\beta}}\end{bmatrix}\boldsymbol{d}_n + \begin{bmatrix}\boldsymbol{\Sigma}_\alpha \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\nu}_{n\alpha} & \boldsymbol{\nu}_{\boldsymbol{n\beta}}\end{bmatrix} \tag{4.2}$$

- and where

$$\boldsymbol{d}_n \sim F_{\boldsymbol{d}}(\boldsymbol{d}) \qquad \boldsymbol{\nu}_n \sim F_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \tag{4.3}$$

  - note that the person specific coefficients are equal to the mean value of the parameters $\boldsymbol{\theta}_1 = \begin{bmatrix}\alpha & \boldsymbol{\beta}'\end{bmatrix}'$, plus deviation from the mean due to a second set of parameters $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$ and given by $\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n$
  - each consumer is assumed to have a fixed set of coefficients $\{\alpha_n, \boldsymbol{\beta_n}\}$
  - we do not impose the restriction that taste parameters $\{\alpha, \boldsymbol{\beta}\}$ – the marginal utilities of product characteristics – are the same for all consumers
  - the person specific coefficients are modeled as a function of underlying common parameters $\{\boldsymbol{\Pi} \text{ and } \boldsymbol{\Sigma}\}$ that are multiplied to the person specific characteristics $(\boldsymbol{d}_n, \boldsymbol{\nu}_n)$, each of which are random draws from an underlying mean zero population with distribution functions $F_{\boldsymbol{d}}(\boldsymbol{d})$ and $F_{\boldsymbol{\nu}}(\boldsymbol{\nu})$

- Let π_{ab} and σ_{ef} be the terms of $\mathbf{\Pi}$ and $\mathbf{\Sigma}$ respectively and let ($\boldsymbol{d}_n = (d_{1n}, \ldots, d_{5n})'$) be the five demographics of the n^{th} person recorded as deviation from the population mean values – then

$$\begin{aligned} \alpha_n = \alpha \quad & +\pi_{11}d_{1n} + \pi_{12}d_{2n} + \ldots + \pi_{15}d_{5n} \\ & +\sigma_{11}v_{1n} + \sigma_{12}v_{2n} + \ldots + \sigma_{14}v_{4n} \\ \beta_{kn} = \beta_k \quad & +\pi_{k1}d_{1n} + \pi_{k2}d_{2n} + \ldots + \pi_{k5}d_{5n} \\ & +\sigma_{k1}v_{1n} + \sigma_{k2}v_{2n} + \ldots + \sigma_{k4}v_{4n} \end{aligned} \tag{4.4}$$

- If there are D person specific observed characteristics ($\boldsymbol{d}_n = (d_{1n}, \ldots, d_{Dn})'$) and $k-1$ product characteristics, then $\mathbf{\Pi}$ is a $k \times D$ and Σ is a $k \times k$ matrix of parameters, i.e.,

$$\underbrace{\begin{bmatrix} \alpha_n \\ \boldsymbol{\beta}_n \end{bmatrix}}_{k\times 1} = \underbrace{\begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix}}_{k\times 1} + \underbrace{\mathbf{\Pi}\boldsymbol{d}_n}_{k\times D \text{ by } D\times 1} + \underbrace{\mathbf{\Sigma}\boldsymbol{\nu}_n}_{k\times k \text{ by } k\times 1} \tag{4.5}$$

  - suppose there are three observed product characteristics (so $k-1=3$)
  - five observed person specific characteristics so that $\begin{bmatrix} \alpha & \boldsymbol{\beta}' \end{bmatrix}'$ is a 4×1 vector (the additional dimension is for price) and $\boldsymbol{d}_n$ is a 5×1 vector
  - $\boldsymbol{\nu}_n$ is also a 4×1 vector – these are the person specific random error terms that provide part of the deviation from the mean values of $\begin{bmatrix} \alpha & \boldsymbol{\beta}' \end{bmatrix}'$
  - Then $\mathbf{\Pi}$ is 4×5 matrix (20 parameters) and $\mathbf{\Sigma}$ is a 4×4 matrix (16 parameters) and so the total number of parameters affecting the utility function are 4 + 20 + 16 = 40

- If we insert (4.2) back into (4.1) and simplify, then the utility function can be decomposed into three parts (or four, if we count $\alpha_n y_n$ term, but it drops out later on)

$$
\begin{aligned}
&u_{njt} = \alpha_n y_n + \delta_{jt} + \mu_{njt} + \epsilon_{njt} \\
&\text{where,} \\
&\delta_{jt} = \delta(\mathbf{x}_{jt}, p_{jt}, \xi_{jt}; \boldsymbol{\theta}_1) = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \\
&\mu_{njt} = \mu(\mathbf{x}_{jt}, p_{jt}, \boldsymbol{d}_n, \boldsymbol{\nu}_n; \boldsymbol{\theta}_2) = (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)
\end{aligned}
\tag{4.6}
$$

- Note the following
  - except for the μ_{njt} term, which arises due to multiplication of $(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)$ with the observed product characteristics, the rest of the form is the same as in the logit case
  - as before, $\alpha_n y_n$ will drop out of the model, δ_{jt} is the mean utility of product j and is common to all consumers
  - $\mu_{njt} + \epsilon_{njt}$ is the mean-zero heteroscedastic error term that captures the deviation from the mean utility
  - it is this last composite error term $\mu_{njt} + \epsilon_{njt}$, that allows us to break away from the IIA property

- Utility can be written as

$$
\begin{aligned}
&u_{njt} = \alpha_n y_n + \delta_{jt} + \mu_{njt} + \epsilon_{njt} \\
&\text{where,} \\
&\delta_{jt} = \delta(\mathbf{x}_{jt}, p_{jt}, \xi_{jt}; \boldsymbol{\theta}_1) = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \\
&\mu_{njt} = \mu(\mathbf{x}_{jt}, p_{jt}, \boldsymbol{d}_n, \boldsymbol{\nu}_n; \boldsymbol{\theta}_2) = (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)
\end{aligned}
\tag{4.6}
$$

- Recall that in the logit model the IIA property was arising due to the independence of the error terms ϵ_{njt}
  - One way around this problem is to allow these error terms to be correlated across different brands – and in principle one can allow a completely unrestricted variance-covariance matrix for the shocks ϵ_{njt} – leads to the dimensionality problem (all pair-wise covariances between products and variances of each of the J products)
  - The nested logit took a restricted version of this by imposing some structure on the error terms so that all products within a group have a correlation between them but not with those in other groups
- In the current context, we retain the iid extreme value distribution assumption on ϵ_{njt}, but the correlation among the choices is generated via the μ_{njt} component of the composite error term $\mu_{njt} + \epsilon_{njt}$
  - Correlation between utility of different products is a function of both product and consumer attributes so that products with similar characteristics will have similar rankings and consumers with similar demographics will have also have similar rankings of products ($\mu_{njt} = (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)$)
  - Rather than estimate a large number of parameters of a completely unrestricted variance-covariance matrix for ϵ_{njt}, we need to estimate relatively fewer parameters $\boldsymbol{\theta}_1 = (\alpha,\ \boldsymbol{\beta})'$, $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$

- Utility of product j for two different consumers differs only by $\mu_{njt} + \epsilon_{njt}$ (see (4.6) – $u_{njt} = \alpha_n y_n + \delta_{jt} + \mu_{njt} + \epsilon_{njt}$)
  - the δ_j term is the same for all consumers and $\alpha_n y_n$ is the same for all choices
  - hence the fact that one consumer choose product j while another chooses product i must only be because the two consumers differ in their product specific idiosyncratic error terms $\mu_{njt} + \epsilon_{njt}$
- Hence, we can describe each consumer as a tuple of demographic and product specific shocks $(\boldsymbol{d}_n, \boldsymbol{\nu}_n, \epsilon_{n0t}, \epsilon_{n1t}, \ldots, \epsilon_{nJt})$, which implicity defines the set of individual attributes that choose product j given by

$$\mathbb{A}_{jt}(\mathbf{x}_t, \mathbf{p}_t, \boldsymbol{\delta}_t(\mathbf{x}_t, \mathbf{p}_t; \boldsymbol{\theta}_1); \boldsymbol{\theta}_2) = \{(\boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{n0t}, \epsilon_{n1t}, \ldots, \epsilon_{nJt}) \mid u_{njt} > u_{nlt} \quad \forall\, l = 0, 1, 2 \ldots J,\ l \neq j\}. \tag{4.7}$$

- The market share of product j is the integral of the joint distribution of $(\boldsymbol{d}, \boldsymbol{\nu}, \boldsymbol{\epsilon})$ over the mass of individuals in the region A_{jt},

$$s_{jt} = \int_{\mathbb{A}_{jt}} dF(\boldsymbol{d}, \boldsymbol{\nu}, \boldsymbol{\epsilon}) = \int_{\mathbb{A}_{jt}} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) dF_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) \tag{4.8}$$

  - where the second part follows only if we assume that the three random variables for a given consumer are independently distributed
  - note also that set $\mathbb{A}_{jt}$ is only defined via the parameters $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$, since they were part of the μ_{njt} term, and not over the parameters $\boldsymbol{\theta}_1$

- Unlike the logit case, the integral does not have a closed form
- If we continue to assume that ϵ_{njt} has iid extreme value distribution, then the probability that a given individual $\tilde{n}$ – with endowed values of $\tilde{\boldsymbol{d}}_n$ and $\tilde{\boldsymbol{\nu}}_n$, or equivalently with a given value of $\tilde{\mu}_{njt}$ – chooses product j, continues to have a closed logit form like equation 5.6 and in this case is given by

$$s_{njt} = \frac{\exp(\delta_{jt} + \tilde{\mu}_{njt})}{\sum_{j=0}^{J} \exp(\delta_{jt} + \tilde{\mu}_{njt})} \tag{4.9}$$

- Since $\mu_{njt} = \mu(\mathbf{x}_{jt}, p_{jt}, \boldsymbol{d}_n, \boldsymbol{\nu}_n; \boldsymbol{\theta}_2)$, we can integrate individual probability over the distribution of $\boldsymbol{d}_n$ and $\boldsymbol{\nu}_n$ to recover market share of product j

$$\begin{aligned} s_{jt} &= \int_{\mathbb{A}_{jt}} s_{njt} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \\ &= \int_{\mathbb{A}_{jt}} \left\{ \frac{\exp(\delta_{jt} + \mu_{njt})}{\sum_{j=0}^{J} \exp(\delta_{jt} + \mu_{njt})} \right\} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \end{aligned} \tag{4.10}$$

- Price elasticities of market shares are given by

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int_{\mathbb{A}_{jt}} \alpha_n s_{njt}(1 - s_{njt}) dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) & \text{if } j = k, \\ \frac{p_{kt}}{s_{jt}} \int_{\mathbb{A}_{jt}} \alpha_n s_{njt} s_{nkt} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) & \text{otherwise} \end{cases} \tag{4.11}$$

where $s_{njt} = \frac{\exp(\delta_{jt} + \tilde{\mu}_{njt})}{\sum_{j=0}^{J} \exp(\delta_{jt} + \tilde{\mu}_{njt})}$

- Main advantages of this model are that estimation requires estimation of a handful of parameters (rather than square of the number of parameters), elasticities do not exhibit the problems noted earlier for the logit (own or cross-elasticities) and allows us to model consumer heterogeneity rather than rely on a representative consumer
- Compare to the earlier elasticities from the logit model

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k, \\ \alpha p_{kt} s_{kt} & \text{otherwise} \end{cases} \tag{3.11}$$

- Nothing comes for free ... now we must integrate the expression numerically

- Let x be some arbitrary random variable[2] with a probability distribution $f(x) = dF(x)/dx \rightarrow dF(x) = f(x)dx$
    - then note that the integral – $\int x \cdot f(x)dx$ – is just the expected value of x, i.e., $\text{E}[x] = \int x \cdot dF(x)$
    - the sample analog would be the weighted average of x given by $\bar{x} = \sum_n x_n Pr(x_n)$
    - further, if all values are equally possible, then it is just the simple sample average $\bar{x} = (1/N) \sum_n x_n$
- The idea carries over to any function $g(x)$ defined over x such that
    - $\text{E}[g(x)] = \int g(x) \cdot dF(x)$
    - and the sample analog would be $\overline{g(x)} = \sum_n g(x_n) Pr(x_n)$
- Thus, if we wanted to numerically evaluate the integral of $g(x)$ with a known distribution of x (i.e., evaluate $\int g(x) \cdot dF(x)$), all we need to do is
    - take lots of draws of x from this known distribution
    - evaluate $g(x)$ at each of these points
    - and then just take a simple average of all these values of $g(x)$
    - we will get a pretty good value of the integral by this method if we have taken enough *good* draws of the random variable x

[2] This x has nothing to do with the earlier characteristic vector $\mathbf{x}_{jt}$

- Consider the case where x is distributed between 0 and 3 such that the probabilities of draws are
  - $\Pr(0 \leq x < 1) = .45$,
  - $\Pr(1 \leq x < 2) = .10$, and
  - $\Pr(2 \leq x < 3) = .45$
- If we drew 100 random numbers from this distribution, we would expect about 45 of them to be between 0 and 1, another 10 observations between 1 and 2, and 45 observations between 2 and 3
  - If that were the case, we could safely evaluate $g(x)$ at each of these 100 random draws and take their average to compute $\mathrm{E}[g(x)] = \int g(x) \cdot dF(x)$
  - If on the other hand we find that the drawing sequence (algorithm) is such that for the first 100 draws, we have 1/3 of observations from each of the three regions, then with just 100 draws, average values of $g(x)$ will obviously give a very poor (if not outright wrong) approximation to the integral in question
- There is a large literature on drawing from different types of random distributions, for a good review of basic techniques see chapter 9 in Train

- To compute the integral in (4.10), we need to know the distribution functions $F_{\boldsymbol{d}}(\boldsymbol{d})$ and $F_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ and draw from these distributions
- Drawing from $F_{\boldsymbol{d}}(\boldsymbol{d})$
  - note that $\boldsymbol{d}_n$ is the vector of demographics for consumer n (income, family size, age, gender, etc.)
  - one way to proceed is to make use other of data sources, such as the census data, to construct a non-parametric distribution. We can then take random draws from this distribution to compute the integral above
  - in practice one can directly draw N number of consumers – where N is a reasonably large number – from each of the t markets and record their demographic information
  - thus, let us assume that $\boldsymbol{d}_n$ is a 5×1 vector of demographics, and that we have obtained N_s random draws from each market and recorded the values of these demographics
- Drawing from $F_{\boldsymbol{\nu}}(\boldsymbol{\nu})$
  - recall that if $\mathbf{x}_{jt}$ is a vector of three observed characteristics $(k - 1 = 3)$ for product j, then for each person, $\boldsymbol{\nu}_n$ is a 4×1 (or more generally $k \times 1$) vector of random error terms that provide part of the deviation from the mean values of $\begin{bmatrix} \alpha & \boldsymbol{\beta}' \end{bmatrix}'$
  - researchers often specify $F_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ as standard multivariate normal and take N draws per market to obtain $\boldsymbol{\nu}_n$
  - let us again assume that with the help of a good random number generater, we have taken N_s such draws per market and have recorded a series of 4×1 vectors for each person

- Given the values of the parameters $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$, a value of mean utility δ_{jt} and N_s random values of $\boldsymbol{d}_n$ and $\boldsymbol{\nu}_n$, the predicted market share of good j can be computed using the smooth simulator as the average value of s_{njt} over the N_s observations,

$$\begin{aligned}
\tilde{s}_{jt} &= \int_{A_{jt}} s_{njt} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \\
&= \frac{1}{N_s} \sum_{n}^{N_s} s_{njt} = \frac{1}{N_s} \sum_{n}^{N_s} \left\{ \frac{\exp(\delta_{jt} + \mu_{njt})}{\sum_{j=0}^{J} \exp(\delta_{jt} + \mu_{njt})} \right\}
\end{aligned} \tag{4.12}$$

where $\mu_{njt} = (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)$

- Recall from earlier example (5 demographics and 3+1 product characteristics), there were 40 parameters to estimate
- Data may not allow such estimation of such rich set of parameters
  - BLP do not use individual demographics to create variation in person specific coefficients
  - equivalently, the $k \times d$ matrix $\mathbf{\Pi}$ consists of zeros and the variation in $\begin{bmatrix} \alpha_n & \boldsymbol{\beta}_n' \end{bmatrix}'$ is only due to $\mathbf{\Sigma}\boldsymbol{\nu}_n$
  - Nevo sets only some of the terms of $\mathbf{\Pi}$ to zero and estimates the other coefficients
  - Often researchers set $\mathbf{\Sigma}$ as a diagonal matrix and estimate only the leading terms of this matrix
  - this is not as restrictive as it may appear at first pass

- To understand the logic of choosing parameters that are set to zero, and the implications, consider a very simple example where there is only one observed characteristic of each product, plus price, so that $\begin{bmatrix}\alpha_n & \boldsymbol{\beta}_n'\end{bmatrix}'$ is just a 2×1 column vector instead of $k \times 1$
  - just to be clear, in what follows in the next couple of paragraphs, think of $\boldsymbol{\beta}_n$ and $\boldsymbol{\beta}$ as just 1×1 scalars even though I continue to write them in bold font for vectors
  - Further, suppose that all the elements of $\boldsymbol{\Pi}$ are zero (again, only to simplify the algebra as the main idea carries through with or without $\boldsymbol{\Pi}$ in the utility function)
- Then sans the $\boldsymbol{\Pi d}_n$ term

$$\begin{bmatrix}\alpha_n \\ \boldsymbol{\beta}_n\end{bmatrix} = \begin{bmatrix}\alpha \\ \boldsymbol{\beta}\end{bmatrix} + \boldsymbol{\Sigma\nu}_n = \begin{bmatrix}\alpha \\ \boldsymbol{\beta}\end{bmatrix} + \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22}\end{bmatrix}\begin{bmatrix}\nu_{1n} \\ \nu_{2n}\end{bmatrix} \tag{4.13}$$

- Since $\boldsymbol{\nu}_n$ is a mean zero error term, then

$$\begin{aligned}
\alpha_n &= \alpha + \sigma_{11}\nu_{1n} + \sigma_{12}\nu_{2n} \\
\boldsymbol{\beta}_n &= \boldsymbol{\beta} + \sigma_{21}\nu_{1n} + \sigma_{22}\nu_{2n} \\
\mathrm{E}[\alpha_n] &= \alpha \qquad \mathrm{E}[\boldsymbol{\beta}_n] = \boldsymbol{\beta} \\
\mathrm{Var}[\alpha_n] &= \sigma_{11}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{11}\sigma_{12}\mathrm{Cov}[\nu_{1n}, \nu_{2n}] + \sigma_{12}^2\mathrm{Var}[\nu_{2n}] \\
\mathrm{Var}[\boldsymbol{\beta_n}] &= \sigma_{21}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{21}\sigma_{22}\mathrm{Cov}[\nu_{1n}, \nu_{2n}] + \sigma_{22}^2\mathrm{Var}[\nu_{2n}]
\end{aligned} \tag{4.14}$$

- Since $\boldsymbol{\nu}_n$ is a mean zero error term, then

$$\begin{aligned}
\alpha_n &= \alpha + \sigma_{11}\nu_{1n} + \sigma_{12}\nu_{2n} \\
\boldsymbol{\beta}_n &= \boldsymbol{\beta} + \sigma_{21}\nu_{1n} + \sigma_{22}\nu_{2n} \\
\mathrm{E}[\alpha_n] &= \alpha \qquad \mathrm{E}[\boldsymbol{\beta}_n] = \boldsymbol{\beta} \\
\mathrm{Var}[\alpha_n] &= \sigma_{11}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{11}\sigma_{12}\mathrm{Cov}[\nu_{1n}, \nu_{2n}] + \sigma_{12}^2\mathrm{Var}[\nu_{2n}] \\
\mathrm{Var}[\boldsymbol{\beta_n}] &= \sigma_{21}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{21}\sigma_{22}\mathrm{Cov}[\nu_{1n}, \nu_{2n}] + \sigma_{22}^2\mathrm{Var}[\nu_{2n}]
\end{aligned} \tag{4.14}$$

- Implications of setting the off-diagonal terms in $\boldsymbol{\Sigma}$ to zero: if $\sigma_{12} = \sigma_{21} = 0$, then
  - α_n is a deviation from the mean value of α and the deviation is determined only by a random shock ν_{1n} multiplied by a coefficient σ_{11}
  - the shock to the marginal utility of the second characteristic ν_{2n}, does not affect the deviation from mean for the first characteristics, i.e., the marginal (dis)utility of price
  - put another way, the unobserved heterogeneity has been modeled such that if price and speed of a computer are the only two characteristics in consideration, and a given person gets a positive shock to the marginal utility of speed (they get more utility from the speed of computer relative to another person), it does not imply that they also get a higher (dis)utility from the price of the computer due to the higher utility from speed
  - the (dis)utility from price is equal to α plus a person specific deviation only for price $\sigma_{11}\nu_{1n}$
  - similarly, variances of α_n and $\boldsymbol{\beta}_n$ depend on the variances of the shocks of these characteristics (e.g. $\mathrm{Var}[\alpha_n] = \sigma_{11}^2\mathrm{Var}[\nu_{1n}]$) but not on the *covariance* of the shocks, even if $\mathrm{Cov}[\nu_{1n}, \nu_{2n}] \neq 0$, since $\sigma_{12} = \sigma_{21} = 0$

- Next, consider the covariance between α_n and $\boldsymbol{\beta}_n$
- Covariance between the two random variables is defined as $\text{Cov}(\alpha_n, \boldsymbol{\beta}_n) = \text{E}\left[\{\alpha_n - \text{E}(\alpha_n)\}\{\boldsymbol{\beta}_n - \text{E}(\boldsymbol{\beta}_n)\}\right]$ hence

$$\begin{aligned}\text{Cov}(\alpha_n, \boldsymbol{\beta}_n) &= \text{E}(\alpha_n \boldsymbol{\beta}_n) - \alpha\boldsymbol{\beta} \\ &= \sigma_{11}\sigma_{21}\text{Var}(\nu_{1n}) + \sigma_{12}\sigma_{22}\text{Var}(\nu_{2n}) \\ &+ \sigma_{11}\sigma_{22}\text{Cov}(\nu_{1n}, \nu_{2n}) + \sigma_{12}\sigma_{21}\text{Cov}(\nu_{1n}, \nu_{2n}) \\ &= \sigma_{11}\sigma_{22}\text{Cov}(\nu_{1n}, \nu_{2n}).\end{aligned} \tag{4.15}$$

- the first line is due to the definition of a covariance and the observation that $\text{E}[\alpha_n] = \alpha$ and $\text{E}[\boldsymbol{\beta}_n] = \boldsymbol{\beta}$
- the second line follows from substituting values of α_n and $\boldsymbol{\beta}_n$ from equation (4.14), taking the expectations, setting $\text{E}[\boldsymbol{\nu}_n] = \mathbf{0}$ and simplifying
- the last line is if we set $\sigma_{12} = \sigma_{21} = 0$ and shows that even after setting the off-diagonals in $\boldsymbol{\Sigma}$ equal to zero, the covariance between the marginal utilities is not necessarily zero – *unless we now further assume that the mean zero error terms $\boldsymbol{\nu}_n$ are not correlated across the characteristics*

- Common to assume that $\boldsymbol{\nu}_n$ are drawn from multivariate standard normal or log normal, i.e., covariances between the error terms are zero as well
  - In the special case where the terms of $\boldsymbol{\Pi}$ are also zero – as in the foregoing discussion – this implies that covariances between marginal utilities will also be zero
- However, if terms of $\boldsymbol{\Pi}$ are not all zero, they will still invoke correlations between the marginal utilities of different characteristics
  - as equation (4.4), reproduced below for this special case of two characteristics and five demographics, shows

$$\begin{aligned} \alpha_n = \alpha \quad & +\pi_{11} d_{1n} + \pi_{12} d_{2n} + \ldots + \pi_{15} d_{5n} \\ & \qquad + \sigma_{11}\nu_{1n} + \sigma_{12}\nu_{2n} \\ \boldsymbol{\beta}_n = \boldsymbol{\beta} \quad & +\pi_{21} d_{1n} + \pi_{22} d_{2n} + \ldots + \pi_{25} d_{5n} \\ & \qquad + \sigma_{21}\nu_{1n} + \sigma_{22}\nu_{2n} \end{aligned} \tag{4.4}$$

  - in this case, the covariance between α_n and $\boldsymbol{\beta_n}$ will be invoked via the π terms and the covariances between the demographic variables, even if we set $\sigma_{12} = \sigma_{21} = 0$ and choose the distribution of $\boldsymbol{\nu}_n$ to be multivariate standard normal
- Thus as mentioned earlier, if we use demographic data and don't set the $\boldsymbol{\Pi}$ to zero (at least not all terms) then setting the off diagonals of $\boldsymbol{\Sigma}$ to zero and drawing $\boldsymbol{\nu}_n$ from multivariate standard normal is not so restrictive

- The essential idea of estimation remains the same as that of two-step estimation outlined in the section on logits
- Briefly,
  - estimate mean utility δ_{jt} and then use it in the second step to estimate the moment functions and find parameters that minimize the value
  - this requires first estimating model predicted market shares via (4.10), equating them to observed market shares, and then inverting the relation and using a contraction mapping to compute δ_{jt}
- We consider each of these along the way and following Nevo (2001), combine everything in a 5-step algorithm

Random Coefficients Logit
Estimation Details

(-1) For each market t, draw N_s random values for $(\boldsymbol{\nu}_n, \boldsymbol{d}_n)$ from the distributions $F_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ and $F_{\boldsymbol{d}}(\boldsymbol{d})$

- the distribution $F_{\boldsymbol{d}}(\boldsymbol{d})$ can be estimated using census data
- for $F_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ we can use zero mean multivariate normal with a pre-specified covariance matrix

(0) Select arbitrary initial values of δ_{jt} and $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$ and for $\boldsymbol{\theta}_1$

- for $\boldsymbol{\theta}_1 = \begin{bmatrix}\alpha & \boldsymbol{\beta}'\end{bmatrix}'$ use initial values from simple logit estimation

(1) Use random draws and the initial parameter values to estimate model predicted market shares $\tilde{s}_{jt}$ of each product in each market

- use (4.12) to compute these shares

(2) Obtain $\widehat{\delta}_{jt}$

(a) Keep $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$ fixed and change values of δ_{jt} until predicted shares $\tilde{s}_{jt}$ in step above, equal the observed shares – this is the inversion step where we want to find $\boldsymbol{\delta}_t$ such that $s_{jt} = \tilde{s}_{jt}(\delta_{1t}, \ldots, \delta_{Jt}, \boldsymbol{\theta}_2)$ in each market

(b) This can be done using the contraction mapping $\boldsymbol{\delta}_t^{h+1} = \boldsymbol{\delta}_t^h + [ln(\mathbf{s}_t) - ln(\tilde{\mathbf{s}}_t)]$

(c) Note carefully that mean utility is a function of observed market shares and parameters $\boldsymbol{\theta}_2$ thus, $\delta_{jt} = \delta_{jt}(\mathbf{s}_t, \boldsymbol{\theta}_2)$

(3) Define error term as $\xi_{jt} = \widehat{\delta}_{jt}(\mathbf{s}_t, \boldsymbol{\theta}_2) + \alpha p_{jt} - \mathbf{x}_{jt}\boldsymbol{\beta}$ and calculate the value of the moment condition, i.e., the GMM objective function

- (A) As before, subsume p_{jt} within $\mathbf{x}_{jt}$ as just another column of $\mathbf{x}_{jt}$ and redefine $\mathbf{x}_{jt} = \begin{bmatrix} -p_{jt} & \mathbf{x}_{jt} \end{bmatrix}$; similarly, redefine matrix $\mathbf{X}$ to be inclusive of the price vector so that $\mathbf{X} = \begin{bmatrix} -\mathbf{p} & \mathbf{X} \end{bmatrix}$
- (B) Thus $\xi_{jt}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \widehat{\delta}_{jt}(\mathbf{s}_t, \boldsymbol{\theta}_2) - \mathbf{x}_{jt}\boldsymbol{\theta}_1$.
  In matrix notation $\boldsymbol{\xi} = \widehat{\boldsymbol{\delta}}(\mathbf{s}, \boldsymbol{\theta}_2) - \mathbf{X}\boldsymbol{\theta}_1$
- (C) Then the objective function to be minimized is
  $\left(\boldsymbol{\xi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'\mathbf{Z}\right)\boldsymbol{\Phi}\left(\mathbf{Z}'\boldsymbol{\xi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\right)$,
  where $\boldsymbol{\Phi}$ is the GMM weighting matrix
- (D) Initially set the weighting matrix as $\boldsymbol{\Phi} = (\mathbf{Z}'\mathbf{Z})^{-1}$

(4) Search for better values of $\boldsymbol{\theta}_1 = \begin{bmatrix} \alpha & \boldsymbol{\beta}' \end{bmatrix}'$ and $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$ and the GMM weighting matrix $\boldsymbol{\Phi}$ that minimize the objective function as follows:

(A) Note that while $\boldsymbol{\xi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is a function of both sets of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, it actually partitions into two components: $\xi_{jt}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \widehat{\delta}_{jt}(\mathbf{s}_t, \boldsymbol{\theta}_2) - \mathbf{x}_{jt}\boldsymbol{\theta}_1$
– this is important because we can help the search algorithm by solving for $\boldsymbol{\theta}_1$, conditional on $\boldsymbol{\theta}_2$ analytically – how? in the GMM objective function given above $[(\boldsymbol{\xi}'\mathbf{Z})\boldsymbol{\Phi}(\mathbf{Z}'\boldsymbol{\xi})]$, set $\boldsymbol{\xi} = \widehat{\boldsymbol{\delta}}(\boldsymbol{\theta}_2) - \mathbf{X}\boldsymbol{\theta}_1$
– now consider the first order condition with respect to $\boldsymbol{\theta}_1$ and solve for $\boldsymbol{\theta}_1$. See equations 5.31 and 5.32 for FOC and its solution for the GMM estimator
– this implies that if we have some fixed values of $\boldsymbol{\theta}_2$, then $\boldsymbol{\theta}_1$ can be solved for analytically as $\boldsymbol{\theta}_1 = (\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\widehat{\boldsymbol{\delta}}(\boldsymbol{\theta}_2)$

(B) Thus, first solve (search) for $\boldsymbol{\theta}_1$ as $\widehat{\boldsymbol{\theta}}_1 = (\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\widehat{\boldsymbol{\delta}}(\boldsymbol{\theta}_2)$

(C) Use new $\boldsymbol{\theta}_1 = \begin{bmatrix} \alpha & \boldsymbol{\beta}' \end{bmatrix}'$ to re-compute error term $\boldsymbol{\xi}$ (see 3b above)

(D) Next, update the weighting matrix $\boldsymbol{\Phi}$ as $\boldsymbol{\Phi} = (\mathbf{Z}'\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{Z})^{-1}$

(E) Take the new value of $\boldsymbol{\Phi}$ and update the GMM objective function, $(\boldsymbol{\xi}'\mathbf{Z})\boldsymbol{\Phi}(\mathbf{Z}'\boldsymbol{\xi})$

(F) Finally, update $\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$ – do a non-linear search over $\{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$ to minimize the objective function

(5) Return to step (1) above with all new shiny parameter values (keep the original draws) and iterate
   – Note that you can skip the updating of the weighting matrix $\mathbf{\Phi}$ in step 4e from now on

- Brand Dummies
  - In the section on logits, we discussed adding in the brand dummies to the vector $\mathbf{x}_{jt}$ and recovering the $\boldsymbol{\beta}$ coefficients for the brand characteristics
  - Same can be done here as well, but will need to have two separate versions of data matrix $\mathbf{X}$ (call them $\mathbf{X}_1$ and $\mathbf{X}_2$)
  - Observe that $\mathbf{X}$ (defined to be inclusive of the price vector) enters the utility function twice:
    - in the linear part of the estimation as mean utility $\boldsymbol{\delta}(\mathbf{X};\boldsymbol{\theta}_1) = \mathbf{X}\boldsymbol{\theta}_1 + \boldsymbol{\xi}$ – this is from $\delta_{jt} = \delta(\mathbf{x}_{jt}, p_{jt}, \xi_{jt}; \boldsymbol{\theta}_1) = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$
    - and in the non-linear part of the estimation as individual deviation from the mean utility $\boldsymbol{\mu}_n(\mathbf{X};\boldsymbol{\theta}_2, \boldsymbol{d}_n, \boldsymbol{\nu}_n) = \mathbf{X}(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)$ – this follows from $\mu_{njt} = (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)$ – and allows for random coefficients on product characteristics
    - In practice we may not want to allow random coefficients on all characteristics, in which case the data matrix $\mathbf{X}$ appearing in $\boldsymbol{\mu}_n$ can be a subset of the one appearing the linear part $\boldsymbol{\delta}$
  - Thus, we can write the two components as
    $\boldsymbol{\delta}(\mathbf{X}_1;\boldsymbol{\theta}_1) = \mathbf{X}_1\boldsymbol{\theta}_1 + \boldsymbol{\xi}$ and,
    $\boldsymbol{\mu}_n(\mathbf{X}_2;\boldsymbol{\theta}_2, \boldsymbol{d}_n, \boldsymbol{\nu}_n) = \mathbf{X}_2(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)$

- Brand Dummies
  - Thus, we can write the two components as $\boldsymbol{\delta}(\mathbf{X}_1;\boldsymbol{\theta}_1) = \mathbf{X}_1\boldsymbol{\theta}_\mathbf{1} + \boldsymbol{\xi}$ and, $\boldsymbol{\mu}_n(\mathbf{X}_2;\boldsymbol{\theta}_2,\boldsymbol{d}_n,\boldsymbol{\nu}_n) = \mathbf{X}_2(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)$
  - $\mathbf{X}_1$ includes all variables that are common to all individuals (price, promotional activities, and brand characteristics or brand dummies instead of brand characteristics)
  - $\mathbf{X}_2$ contains variables that can have random coefficients (price and product characteristics but not brand dummies)
  - Note that if we use $\mathbf{X}_1$ and $\mathbf{X}_2$, then the estimator $\widehat{\boldsymbol{\theta}}_1$ in step 4a/4b above will be $\widehat{\boldsymbol{\theta}}_1 = (\mathbf{X}_1'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\widehat{\boldsymbol{\delta}}(\boldsymbol{\theta}_2)$

- Additional Instruments
  - The instruments matrix $\mathbf{Z}$ consists of all exogenous variables
  - If the brand characteristics (excluding price) are exogenous, then the brand characteristics plus the instrument(s) for the price variable consist of the matrix $\mathbf{Z}$, or alternatively, if we use brand dummies, then the brand dummies and the price instrument(s) form the matrix $\mathbf{Z}$
  - However, note that if we have only one additional instrument for price, it will not be enough for identification of the model parameters
    - The brand characteristics (or brand dummies) plus the one additional instrument for price will give *exactly* as many moment conditions as the number of components of the parameter vector $\boldsymbol{\theta}_1$
    - These would be enough in the linear logit case
    - However, in the random coefficients case, we have to estimate additional $k \times D + k \times k$ parameters of $\boldsymbol{\theta}_2 = \{\mathbf{\Pi}, \mathbf{\Sigma}\}$
    - This is not possible unless we have additional $k \times D + k \times k$ moment conditions
    - In practice, researchers often set some of the terms of the $\mathbf{\Pi}$ matrix to zero and also set the parameter matrix $\mathbf{\Sigma}$ to be diagonal (see earlier discussions)
    - This reduces the need for additional moment conditions from $kD + k^2$ to $g + k$ where g is the number of non-zero terms in $\mathbf{\Pi}$

- Additional Instruments
  - These may be relatively easier to overcome (these instruments should also not be nearly collinear else will give rise to redundant moment conditions)
  - If one is using BLP style instruments for price (and product characteristics are exogenous) then recall that, in general, one gets more than one instrument for price by using sums of the values of characteristics of other products offered by a firm, and the sums of the values of the same characteristics of products offered by other firms
  - Alternatively, if using Hausman style instruments, the price of the product from more than one market needs to be used (for instance, Nevo (2001) uses data from 20 quarters and multiple cities and constructs 20 additional instruments from other cities matching one from each quarter)
  - An additional set of instruments could be the average value (average over n individuals) of the product characteristics interacted with the person specific demographics to account for the parameters in the $\mathbf{\Pi}$ matrix and similarly the average value of the person specific shocks $\boldsymbol{\nu}$ interacted with product characteristics

Summary

- These lectures are a starting point – they are by no means complete in the sense of covering all the important variants of the models discussed above
- Important variants include
  - using individual level data (in addition to the aggregate data)
  - adding in the cost side moment restrictions to the model (e.g. $\mathbf{p} = \mathbf{c} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}(\boldsymbol{p}, \boldsymbol{z}; \boldsymbol{\xi}))$
  - and modeling dynamic demand (a complicated system of estimation)
- Canned routines in SAS, STATA etc. can be used for linear models (multi-budgeting with AIDs, logit, nested logit etc.)
- For random coefficients models, no canned routine exists (yet) but several researchers have helpfully provided copies of their own code which can serve as very good starting point for coding your own work
- Hopefully, these notes should help in getting started

References

Ackerberg, D., Benkard, C. L., Berry, S., and Pakes, A. (2007). Econometric tools for analyzing market outcomes. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6A, chapter 63, page 41714276. Elsevier.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.

Bokhari, F. A. S. and Fournier, G. M. (2013). Entry in the ADHD drugs market: Welfare impact of generics and me-toos. *Journal of Industrial Economics*, 61(2):340–393.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge.

Deaton, A. and Muellbauer, J. (1980). *Economics and consumer behavior*. Cambridge University Press, Cambridge, UK.

Hausman, J. A., Leonard, G., and Zona, J. (1994). Competitive analysis with differentiated products. *Annales d'Economie et de Statistique*, 34:159–180.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy*, 9(4):513–548.

References

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

Reiss, P. C. and Wolak, F. A. (2007). Structural econometric modeling: rationales and examples from industrial organization. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6A, chapter 64, pages 4277–4415. Elsevier.

Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge University Press, Cambridge.

- If there are L firms and firm l produces a subset $\mathfrak{L}_l$ of all the products, then it maximizes its joint profit over these products as

$$\Pi_l = \sum_{j \in \mathfrak{L}_l} (p_j - mc_j) q_j(\mathbf{p}, \mathbf{z}_j, \xi_j), \tag{5.1}$$

  where c_j is the constant marginal cost of the j-th product, and where $q_j(\cdot) = M s_j(\cdot)$.
- Using Bertrand-Nash equilibrium in prices, the implied marginal costs are given by

$$\mathbf{p} = \mathbf{c} + \boldsymbol{\Omega}^{-1} \boldsymbol{s(p)}$$
$$\Rightarrow$$
$$\mathbf{c} = \mathbf{p} - \boldsymbol{\Omega}^{-1} \boldsymbol{s(p)}$$

- where
  - $\boldsymbol{\Omega}$ is defined such that $\Omega_{jr} = -\Theta_{jr} \frac{\partial s_r(\mathbf{p})}{\partial p_j}$
  - $\boldsymbol{\Theta}$ is 1/0 joint "ownership" matrix with ones in the leading diagonals and in r, j position if these products are produced by the same firm and zeros everywhere else

Merger Simulation

- Let $\boldsymbol{\Theta}_o$ the pre-merger ownership matrix and the related markup matrix be $\boldsymbol{\Omega}_o^{-1}$.
- Suppose two firm merge. The the ownership matrix and the markup matrix will change. Then post merger, let these be given by $\boldsymbol{\Theta}_n$ and $\boldsymbol{\Omega}_n^{-1}$ respectively.
- The the predicted post merger equilibrium prices, $\mathbf{p}_n$, at old marginal costs will be given by

$$\mathbf{p}_n = \mathbf{c} + \boldsymbol{\Omega}_{\boldsymbol{n}}^{-1}(\boldsymbol{p}_{\boldsymbol{n}})\boldsymbol{s}(\boldsymbol{p}_{\boldsymbol{n}})$$

- where $\boldsymbol{\Omega}_n$ is defined such that $\Omega_{jrn} = -\Theta_{jrn}\frac{\partial s_r(\mathbf{p_n})}{\partial p_{nj}}$
- Three Steps
  - Step 1: Obtain marginal cost using $\boldsymbol{\Omega}_o^{-1}$ as $\mathbf{c} = \mathbf{p}_o - \boldsymbol{\Omega}_{\boldsymbol{o}}^{-1}\boldsymbol{s}(\boldsymbol{p}_{\boldsymbol{o}})$
  - Step 2: Construct the new ownership matrix $\boldsymbol{\Theta}_n$
  - Step 3: Search for a new equilibrium price $\mathbf{p}_n$ such that $|\mathbf{p}_{h+1} - \mathbf{p}_h| < \epsilon$ and where $\mathbf{p}_{h+1} = \mathbf{c} + \boldsymbol{\Omega}_{\boldsymbol{n}}^{-1}(\boldsymbol{p}_{\boldsymbol{h}})\boldsymbol{s}(\boldsymbol{p}_{\boldsymbol{h}})$