Advanced Econometric Topics

Methods of Demand Estimation

School of Economics
University of East Anglia

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

Fall 2015

- **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: (1) estimate demand systems using aggregate sales data on a set of differentiated products via a number of techniques including, nested AIDS and random coefficients logit models; and (2) master general programming in SAS and MATLAB.

- **Schedule:** Lectures will take place over a span of 5 weeks ($\sim$ 2 hrs per lecture with a short break in between). The schedule is as given below. The lectures will from 3-5 pm on November 6, 13, 20, 27 and December 4 (all Friday afternoons).
- **SAS Lecture:** In addition to these you are encouraged to register and attend my introduction to SAS session as some of the exercises in this module will use SAS code. The Introductory SAS class is on Monday November 16 from 10-12 (PPD Lecture title: Using SAS for Statistical Analysis).

- **Course Content and Outline:** We will develop a general framework for structural econometrics, but primarily focus on differentiated products competition. Demand estimation forms the bedrock of much of research, and this short course will look at several papers related to the estimation problem and application of estimates to economic questions. There will be an emphasis on empirical (data-related) work, although theoretical work will also be presented and discussed. We will consider both, product space and characteristics space approaches to demand estimation. Topics covered will be AIDS models, logit estimation, overview of the BLP method, and estimation of nested and random coefficients logit models. The class presentations will be a mixture of lectures, discussions of empirical papers, some of which will be student led, and demonstration of computer code using SAS and MATLAB for estimating these 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 Product Space
    1. Homotheticity, Gorman Polar Forms and Aggregation
    2. Separability and Multistage Budgeting
    3. AIDS Model
    4. Estimation Details
  - 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. Some of these are required readings and are listed below.
- General Background Readings
  - 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.
  - Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*.
    Cambridge University Press, Cambridge – Chapter 6.
  - Train, K. E. (2003). *Discrete choice methods with simulation*.
    Cambridge University Press, Cambridge – Chapters 3 & 9.
  - Deaton, A. and Muellbauer, J. (1980b). *Economics and consumer behavior*.
    Cambridge University Press, Cambridge, UK – Chapters 3 & 5.

- **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. Some of these are required readings and are listed below.
- Required Readings

  - Deaton, A. and Muellbauer, J. (1980a). An almost ideal demand system. *American Economic Review*, 70(3):312–326.
  - Hausman, J. A., Leonard, G., and Zona, J. (1994). Competitive analysis with differentiated products. *Annales d'Economie et de Statistique*, 34:159–180.
  - Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.
  - 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.

- **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. Some of these are required readings and are listed below.
- Further Readings for Discussions/Presentations (Student Led)

  - Ellison, S. F., Cockburn, I., Griliches, Z., and Hausman, J. A. (1997). Characteristics of demand for pharmaceutical products: an examination of four cephalosporins. *RAND Journal of Economics*, 28(3):426–446.
  - Chaudhuri, S., Goldberg, P. K., and Jia, P. (2006). Estimating the effects of global patent protection in pharmaceuticals: a case study of quinolones in India. *American Economic Review*, 96(5):1477–1514.
  - Bokhari, F. A. 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.
  - Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729.

- Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium.
  Econometrica, 63(4):841–890.
- Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: the case of the U.S. automobile industry.
  Econometrica, 63(4):pp. 891–951.
- Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry.
  Econometrica, 69(2):307–342.

Module Information
Acknowledgements

These lecture notes are based on a number of sources and draw heavily from the following articles/chapters: **?**, Chap. 6; **?**, Chap. 3 & 5; **?**; **?**; **?**; **?**; **?**; **??**. In addition to these primary sources, I have also benefitted from presentations/lecture notes on the same topics by other researchers who have generously put their slides on the internet. These sources include (1) Matthew Shum (Lecture notes: Demand in differentiated-product markets); (2) Matthijs Wildenbeest (Structural Econometric Modelling in Industrial Organization); (3) Eric Rasmusen (The BLP Method of Demand Curve Estimation in Industrial Organization); (4) John Asker and Allan Collard-Wexler (Demand Systems for Empirical Work in IO); (5) Jonathan Levin (Differentiated Products Demand Systems); (6) Ariel Pakes (NBERMetrics); and (7) Aviv Nevo (NBER Methods Lecture – Estimation of Static Discrete Choice Models Using Market Level Data). Finally, I am also in debt to my colleague Franco Mariuzzo for providing significant feedback on these notes. All errors are mine.

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

$$\frac{p - mc}{p} \tag{1.1}$$

- But cost is 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

- 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

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

or equivalently,

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

- Thus, if we can estimate demand elasticity, we can back out the markups
- Idea extends to oligopoly as well

- When there are differentiated products, we want to estimate the system of demand equations and infer the markups using the full cross-elasticity matrix
- These estimates can then be used in a variety of different contexts, including
  - merger simulations to predict post merger prices
  - estimating the value of new goods (via changes in consumer surplus),
  - other policy questions such as allowing direct to consumer advertising or parallel trade for pharmaceutical products etc.
- The process thus begins with estimating a system of demand equations
- 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 AIDS model but within the context of multistage budgeting as well as variants of logit models derived from random utility/discrete choice models

- Common Problems
  - endogeneity
  - multicollinearity
  - the dimensionality problem
  - unobserved heterogeneity among consumers
- Topology of Various Approaches
  - single vs multi-products
  - product or characteristics space
  - representative vs heterogenous agents
- Depending on the context and the question, a researcher needs to be careful about choosing the appropriate estimation methodology, as there are tradeoffs between how well different methods deal with these issues and how relevant any given problem is within a context

- Should we be measuring demand for aggregate product type (drugs) or individual brands? Prices move together
- Most products have substitutes or complements and it is often necessary to explicitly account for the substitution possibilities to adequately answer the research question at hand
- In the context of multiproducts, the researcher also has to face the problem of dimensionality and multicollinearity
  - 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 $\mathbf{z}$)
  - Restrictions ...

- Imposing the symmetry of the Slutsky matrix or adding up restrictions (Engle and Cournout aggregation) reduces the number of parameters to be estimated.
- However, the essential problem, that the number of parameters increases in the square of the number of products, remains.
  - Slutsky equation: $\frac{\partial q_j}{\partial p_i} = \frac{\partial h_j}{\partial p_i} - q_i \frac{\partial q_j}{\partial y}$
  - Engle aggregation: $\sum_j s_j \eta_{jy} = 1$
  - Cournot aggregation: $\sum_j s_j \eta_{ji} = -s_i$) where η_{ji}
    - q_j and h_j are the marshallian and hicksian demand functions respectively for product j, and y is the income or total expenditure
    - η_{ji} is the cross price elasticity of product j with respect to price of i, η_{jy} is the income elasticity of product j and s_i, s_j are the expenditure shares

- *If the research question allows*, avoid the problem of estimating too many parameters by working with a more restrictive form
- Consider the constant elasticity of substitution (CES) utility function

$$u(\mathbf{q};\rho) = u(q_1, q_2, \ldots, q_J; \rho) = \Big(\sum_i^J q_i^\rho\Big)^{1/\rho} \tag{1.7}$$

  where ρ is the parameter of interest that measures the elasticity of substitution
- The demand for a representative consumer is then given by

$$q_j(\mathbf{p}, I; \rho) = \frac{p_j^{1/(1-\rho)}}{\sum_i^J p_i^{\rho/(1-\rho)}} I \qquad j = 1, \ldots, J. \tag{1.8}$$

- Need to estimate only one parameter ... not J^2 – problem solved!
- But now the cross elasticity between products i and j is the same as between k and j for all combinations of i, j, k,

$$\frac{\partial q_i}{\partial p_j}\frac{p_j}{q_i} = \frac{\partial q_k}{\partial p_j}\frac{p_j}{q_k} \qquad \forall i, j, k. \tag{1.9}$$

- An alternative to the single parameter of the CES utility function is the logit demand (Anderson, de Palma, and Thisse, 1992)

$$u(\mathbf{q};\boldsymbol{\delta}) = \sum_{j}^{J} \delta_j q_j - \sum_{j}^{J} q_j \ln q_j . \tag{1.10}$$

- Elasticities in this model depend on market shares (given by J number of parameters δ_j) but not on the similarities among the products
- What if products j and k are more alike (coke,pepsi) and product i is somewhat more different (fanta)?
- Will discuss logit properties further (later)

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

$$\begin{aligned} q_t^d &= \beta_{10} + \gamma_{12} p_t + \beta_{11} x_{1t} + \xi_{1t}, \\ p_t &= \beta_{20} + \gamma_{22} q_t^s + \beta_{22} x_{2t} + \xi_{2t}, \\ q_t^s &= q_t^d \end{aligned} \tag{1.11}$$

- error terms are such that[1]

$$\begin{aligned} E(\xi_{1t}|\mathbf{x}_t) = 0, E(\xi_{2t}|\mathbf{x}_t) = 0, \\ E(\xi_{1t}^2|\mathbf{x}_t) = \sigma_1^2, E(\xi_{2t}^2|\mathbf{x}_t) = \sigma_2^2 \\ E(\xi_{1t}\mathbf{x}_t) = 0, E(\xi_{2t}\mathbf{x}_t) = 0, \\ \text{and } E(\xi_{1t}\xi_{2t}|\mathbf{x}_t) = 0 \end{aligned} \tag{1.12}$$

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 ...
- solve for the reduced form equilibrium values of q^* and p^* – dropping subscript t, we get

$$\begin{aligned} q^* &= \frac{\beta_{10} + \beta_{20}\gamma_{12}}{1-\gamma_{12}\gamma_{22}} + \frac{\beta_{11}}{1-\gamma_{12}\gamma_{22}} x_1 + \frac{\gamma_{12}\beta_{22}}{1-\gamma_{12}\gamma_{22}} x_2 + \frac{\xi_1 + \gamma_{12}\xi_2}{1-\gamma_{12}\gamma_{22}} \\ p^* &= \frac{\beta_{20} + \beta_{10}\gamma_{22}}{1-\gamma_{12}\gamma_{22}} + \frac{\beta_{11}\gamma_{22}}{1-\gamma_{12}\gamma_{22}} x_1 + \frac{\beta_{22}}{1-\gamma_{12}\gamma_{22}} x_2 + \frac{\gamma_{22}\xi_1 + \xi_2}{1-\gamma_{12}\gamma_{22}} \end{aligned} \tag{1.13}$$

- 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 ...
- Useful to explicitly compute the conditional covariance between p and ξ_1
- Note that conditional on $\mathbf{x}_t$,

$$p^* - E(p^*) = \frac{\gamma_{22}\xi_1 + \xi_2}{1 - \gamma_{12}\gamma_{22}} \quad \text{and } \xi_1 - E(\xi_1) = \xi_1 \tag{1.14}$$

Thus

$$\text{cov}(p, \xi_1) = \frac{\gamma_{22}}{1 - \gamma_{12}\gamma_{22}}\sigma_1^2 + \frac{\text{E}(\xi_1\xi_2)}{1 - \gamma_{12}\gamma_{22}} \tag{1.15}$$

- Even if the error terms across the two equations were uncorrelated ($\text{E}(\xi_{1t}\xi_{2t}|\mathbf{x}_t = 0$), 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 $\text{E}(\xi_{1t}\xi_{2t}|\mathbf{x}_t) = 0$
- For completeness – complete system of equations, i.e., the number of equations are equal to the number of endogenous variables – we also require that $\gamma_{12} \neq 1/\gamma_{22}$.

- we can re-write the system in (1.11) in matrix notation

$$\mathbf{y}_t' = \begin{bmatrix} q_t & p_t \end{bmatrix} \mathbf{x}_t = \begin{bmatrix} 1 & x_{1t} & x_{2t} \end{bmatrix} \boldsymbol{\xi}_t' = \begin{bmatrix} \xi_{1t} & \xi_{2t} \end{bmatrix}$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & -\gamma_{22} \\ -\gamma_{12} & 1 \end{bmatrix} \text{ and, } \mathbf{B} = \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{11} & 0 \\ 0 & \beta_{22} \end{bmatrix} \tag{1.16}$$

then, the system of equations above can be written as

$$\mathbf{y}_t'\boldsymbol{\Gamma} - \mathbf{x}_t\mathbf{B} = \boldsymbol{\xi}_t' \tag{1.17}$$

so that the reduced form equation is

$$\mathbf{y}_t' = \mathbf{x}_t\boldsymbol{\Pi} + \mathbf{v}_t' \quad \text{where } \boldsymbol{\Pi} = \mathbf{B}\boldsymbol{\Gamma}^{-1} \text{ and, } \mathbf{v}_t' = \boldsymbol{\xi}_t'\boldsymbol{\Gamma}^{-1} \tag{1.18}$$

Note that in the equation above we are takeing the inverse of the $\boldsymbol{\Gamma}$ – but the inverse exists if the determinant ($\det(\boldsymbol{\Gamma}) = 1 - \gamma_{12}\gamma_{22}$) is not zero, which goes back to the condition $\gamma_{12} \neq 1/\gamma_{22}$ mentioned above

- The moment restrictions in (1.12) (in general we do not need to impose $E(\xi_{1t}\xi_{2t}|\mathbf{x}_t) = 0$) are

$$\begin{aligned} &\mathrm{E}(\boldsymbol{\xi}_t|\mathbf{x}_t) = \mathbf{0}, && \mathrm{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t'|\mathbf{x}_t) = \boldsymbol{\Sigma} \\ &\mathrm{E}(\mathbf{v}_t|\mathbf{x}_t) = \mathbf{0}, && \mathrm{E}(\mathbf{v}_t\mathbf{v}_t'|\mathbf{x}_t) = \boldsymbol{\Omega} \\ &\text{where } \boldsymbol{\Omega} = (\boldsymbol{\Gamma}^{-1})'\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1}. \end{aligned} \tag{1.19}$$

- Estimation can proceed with IV/2SLS (or 3SLS for joint estimation), where the demand equation is estimated using x_{2t} as the instrument, and supply equation is estimated using x_{1t} as the instrument
- If either $\beta_{22} = 0$ or if data on x_{2t} is not available, demand equation cannot be identified/estimated consistently (vice versa for supply equation)
- Since the x's are exogenous variables, they can serve as instruments
  - x_{2t} are cost shifters – they affect production costs; Correlated with p_t but not with ξ_{1t}, hence use as instruments in demand function
  - x_{1t} are demand shifters – affect willingness-to-pay, but not a firm's production costs; Correlated with q_t but not with ξ_{2t}, hence use as instruments in supply function

- 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)

- Tradeoffs in terms of relative strengths and weaknesses in terms of how they deal with different issues
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - Weak instruments
  - New Goods
  - Cross elasticities

- Tradeoffs in terms of relative strengths and weaknesses in terms of how they deal with different issues
  - Dimensionality of Products
    - For large number of products (say $J = 50$), the product space approach leads to the dimensionality problem mentioned earlier, and may require grouping/nesting these products. 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.
  - Dimensionality of Characteristics
  - Weak instruments
  - New Goods
  - Cross elasticities

- Tradeoffs in terms of relative strengths and weaknesses in terms of how they deal with different issues
  - Dimensionality of Products
  - Dimensionality of Characteristics
    - 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.
  - Weak instruments
  - New Goods
  - Cross elasticities

- Tradeoffs in terms of relative strengths and weaknesses in terms of how they deal with different issues
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - Weak instruments
    - Consider the case of demand equation for the quantity of the j^{th} product. In product space, it would be a function of all other prices, so that $q_j = f(p_1, \ldots, p_J)$ and we have J such equations, one for each product. The estimation is thus in 'wide form' and we need to find instruments for each of these prices. In the characteristics space approach, the fundamental issue is the same, but often the resulting functional form is such that estimation is done in 'long form', i.e., the left hand side of the equation, which is typically shares, is also a function of prices, but appears in the equation as a single variable $s_j = f(p_j)$ and there are J such rows per market. Thus, in the product space, when we do a 2SLS estimation, where the first stage involves regressing each price variable on the exogenous set of instruments, we effectively have to estimate J number of first-stage regressions. By contrast, in the characteristics space, we have to do only one first-stage regression as data on all prices is a single column vector that is regressed on a column vector of instruments. Effectively then, the first stage regressions can exhibit weak instruments property in the product space set up while they may not exhibit such a property in the characteristics space approach.
  - New Goods
  - Cross elasticities

- Tradeoffs in terms of relative strengths and weaknesses in terms of how they deal with different issues
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - Weak instruments
  - 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

- Tradeoffs in terms of relative strengths and weaknesses in terms of how they deal with different issues
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - Weak instruments
  - 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.

- Tradeoffs in terms of relative strengths and weaknesses in terms of how they deal with different issues
  - Dimensionality of Products
  - Dimensionality of Characteristics
  - Weak instruments
  - New Goods
  - Cross elasticities

- Consider the demand function of single product j in market t for a representative consumer, given by

$$q_{jt} = \gamma_j + \alpha_j p_{jt} + \mathbf{x}_{jt}\beta_j + \xi_{jt} \tag{1.22}$$

where $\mathbf{x}_{jt}$ is a vector of product characteristics and ξ_{jt} are the unobserved components of demand

  - Interest is in estimating α_j and demand elasticity
  - Even though product specific intercepts γ_j (willingness to pay) have been included in the model, they are demand shifters, and as such do not change the sensitivity to price depending on the level of income or other demographic characteristics such as family size
  - Micro studies often show that the price coefficient depends in an important way on income/wealth, i.e., lower income people care more about price
  - Consequently, if the income distribution varies across the markets, we should expect the price coefficient to vary across these markets, and we need to find a way to allow for it

- Consider the demand function of single product j in market t for a representative consumer, given by

$$q_{jt} = \gamma_j + \alpha_j p_{jt} + \mathbf{x}_{jt}\beta_j + \xi_{jt} \tag{1.22}$$

  where $\mathbf{x}_{jt}$ is a vector of product characteristics and ξ_{jt} are the unobserved components of demand
  - One could make γ_j to be a function of income, but they are still demand shifters and do not change the sensitivity to price. Similarly, other demographic differences may be important to model as well
  - One could potentially include some ad-hoc interaction terms between average values of demographic variables in market t with price (and other product characteristics) but may not represent demand derived from a consumer's utility maximization problem

- To make it a heterogenous agent model, it is more typical to build a micro model where the parameters that enter the utility function of a consumer – say γ_j and α_j – vary over individuals and are perhaps functions of their demographics

  In that case, the demand equations to be estimated would end up looking something like

$$q_{jt} = \int \gamma_{ij} dG(\gamma_{ij}) + \int \alpha_{ij} p_{jt} dF(\alpha_{ij}) + \mathbf{x}_{jt}\beta_j + \xi_{jt} \tag{1.23}$$

  - where where γ_{ij} and α_{ij} are person and product specific random intercepts and slope coefficients, with known or assumed distribution functions $\gamma_{ij} \sim G(\gamma|\tau)$ and $\alpha_{ij} \sim F(\alpha|\theta)$, and where θ and τ are parameters to be estimated and are functions of demographic variables
  - This is called a random coefficients model

- Separability
- Aggregation

- Separability
  - The main method we will look at in the products space approach is one which solves the dimensionality problem by dividing the products into small sub-groups and then allowing some relatively flexible substitution patterns between the products within a group
  - Useful if we could break down the overall consumer decision problem into separate parts, some of which could be estimated separately
  - This is the issue of **separability**
  - What assumptions do we need on an individual consumer's utility function to treat analyze demand for some products separately from the demand for other products?
- Aggregation

- Separability
- Aggregation
  - A related problem is that of **aggregation**, which considers the relationship between individual consumers' behavior and aggregate consumer behavior (which is the sum of individual behavior over all individuals)
  - When working with aggregate data, one can ask whether there are assumptions on preferences such that aggregate demand is generated by a "representative consumer" with "rationalizeable" preferences
  - There is no reason why aggregate data, or any data that are an average over many people should conform to a theory of consumer behavior that focuses on individual people or households

- Separability
- Aggregation

- Preferences ($\succeq$)are homothetic if $t\mathbf{q}_1 \succeq t\mathbf{q}_2 \Leftrightarrow \mathbf{q}_1 \succeq \mathbf{q}_2$ for any $t > 0$

- Preferences ($\succeq$)are homothetic if $t\mathbf{q}_1 \succeq t\mathbf{q}_2 \Leftrightarrow \mathbf{q}_1 \succeq \mathbf{q}_2$ for any $t > 0$
  - the consumer is indifferent between bundles $t\mathbf{q}_1$ and $t\mathbf{q}_2$ whenever they are indifferent between bundles $\mathbf{q}_1$ and $\mathbf{q}_2$
  - there is only one indifference curve and any indifference curve is a radial blowup of another, and all indifference sets are related by proportional expansion along rays
  - marginal rates of substitution are unaffected by equal proportional changes in all quantities, so that income expansion paths are straight lines through the origin
  - preferences are homothetic if and only if they are of the form

$$u(\mathbf{q}) = F(f(\mathbf{q})) \text{ where } f(t\mathbf{q}) = tf(\mathbf{q}), \tag{2.1}$$

    and $F(\cdot)$ is a monotone increasing function
  - the utility function must admit a function that is homogenous of degree one (the $f(\cdot)$) and since utility functions are only defined upto monotonic transformations, then we may as well write the utility function to be just $u(\mathbf{q}) = f(\mathbf{q})$ where the latter is, as before, homogeneous of degree one

- Consider the consumer's expenditure minimization problem $\min \mathbf{p} \cdot \mathbf{q}$ s.t. $u(\mathbf{q}) = f(\mathbf{q}) = u$.
  - Since the function is homogenous of degree one, doubling $\mathbf{q}$ will double the target utility, but doubling $\mathbf{q}$ means doubling the expenditure
  - This means that if $e(\mathbf{p}, u) = \mathbf{q}^* \cdot \mathbf{p}$ is the minimum expenditure for target utility u, then for a target utility of tu, the minimum expenditure is $e(\mathbf{p}, tu) = t\mathbf{q}^* \cdot \mathbf{p} = te(\mathbf{p}, u)$
  - Now if the initial target utility is equal to 1, then by letting $t = u$, we can write $e(\mathbf{p}, u) = ue(\mathbf{p}, 1)$ and hence, for homothetic utility preferences, the expenditure function is of the form

$$e(\mathbf{p}, u) = ub(\mathbf{p}), \tag{2.2}$$

    where $b(\mathbf{p})$ is some linearly homogenous and concave function of prices

- Consider the consumer's expenditure minimization problem $\min \mathbf{p} \cdot \mathbf{q}$ s.t. $u(\mathbf{q}) = f(\mathbf{q}) = u$.
  - Since the function is homogenous of degree one, doubling $\mathbf{q}$ will double the target utility, but doubling $\mathbf{q}$ means doubling the expenditure
  - This means that if $e(\mathbf{p}, u) = \mathbf{q}^* \cdot \mathbf{p}$ is the minimum expenditure for target utility u, then for a target utility of tu, the minimum expenditure is $e(\mathbf{p}, tu) = t\mathbf{q}^* \cdot \mathbf{p} = te(\mathbf{p}, u)$
  - Now if the initial target utility is equal to 1, then by letting $t = u$, we can write $e(\mathbf{p}, u) = ue(\mathbf{p}, 1)$ and hence, for homothetic utility preferences, the expenditure function is of the form

$$e(\mathbf{p}, u) = ub(\mathbf{p}), \tag{2.2}$$

    where $b(\mathbf{p})$ is some linearly homogenous and concave function of prices
- This implies the following forms for indirect utility, hicksian and marshallian demand curves ($V(\mathbf{p}, y)$, $h(\mathbf{p}, u)$ and $q(\mathbf{p}, y)$ respectively)

$$V(\mathbf{p}, y) = \frac{y}{b(\mathbf{p})}, \qquad h_j(\mathbf{p}, u) = u\frac{\partial b(\mathbf{p})}{\partial p_j}, \qquad q_j(\mathbf{p}, y) = yq_j(\mathbf{p}), \tag{2.3}$$

where $y = \sum_j p_j q_j$ is the total expenditure

- Example: cobb-douglas utility function given by $u(\mathbf{q}) = q_1^{\beta_1} q_2^{\beta_2} \ldots, q_J^{\beta_J}$ where the associated demand functions are of the form

$$q_j = y \frac{1}{p_j} \frac{\beta_i}{\sum_j^J \beta_j}$$

- Implications for demand estimation
  - demand for each good is proportional to expenditure (income), or alternatively, the Engel curve for each good is a straight line going through the origin
  - expenditure elasticity of good j is always one

$$\eta_j = \frac{\partial ln q_j}{\partial \ln y} = 1 \qquad \forall j = 1, \ldots, J.$$

  - known as the expenditure proportionality, which is equivalent to the requirement that budget shares $(w_j = \frac{p_j q_j}{y})$ of all commodities are independent of the level of total expenditure (income) so that a consumer always spends a constant proportion of their income on a product, even though income may be varying across different consumers
  - all expenditure elasticities are equal to one – a result that is contradicted by most empirical studies
  - demand for each good is independent of prices of other products implying that cross-price elasticities are zero

- A less restrictive form is that of **quasi-homotheticity**
- In this formulation, a fixed expenditure element ($a(\mathbf{p})$) is added to the expenditure function in equation (2.2) so that it is now given by

$$e(\mathbf{p}, u) = a(\mathbf{p}) + ub(\mathbf{p}) \tag{2.4}$$

- This form is called the **Gorman Polar Form**
- The term $a(\mathbf{p})$ represents the subsistence level of expenditure when $u = 0$ and $b(\mathbf{p})$ is the marginal cost of utility
- The associated indirect utility and demand functions (per the usual derivations) take the forms

$$\begin{aligned} V(\mathbf{p}, y) &= \frac{y - a(\mathbf{p})}{b(\mathbf{p})} \quad \text{and} \quad q_j(\mathbf{p}, y) = a_j(\mathbf{p}) + \frac{b_j(\mathbf{p})}{b(\mathbf{p})}\Big[y - a(\mathbf{p})\Big] \\ \text{where } a_j(\mathbf{p}) &= \frac{\partial a(\mathbf{p})}{\partial p_j} \quad \text{and} \quad b_j(\mathbf{p}) = \frac{\partial b(\mathbf{p})}{\partial p_j} \end{aligned} \tag{2.5}$$

- $a(\mathbf{p})$ is interpreted as the subsistence spending amount and $b(\mathbf{p})$ is a price index that deflates income/expenditure over and above the subsistence level

- Some authors write it in an alternative form
  - we can define $A(\mathbf{p}) = \frac{1}{b(\mathbf{p})}$ and $B(\mathbf{p}) = -\frac{a(\mathbf{p})}{b(\mathbf{p})}$
  - and define $\alpha_j(\mathbf{p}) = a_j(\mathbf{p}) + b_j(\mathbf{p})B(\mathbf{p}) = a_j(\mathbf{p}) - \beta_j(\mathbf{p})a(\mathbf{p})$ and $\beta_j(\mathbf{p}) = b_j(\mathbf{p})A(\mathbf{p}) = \frac{b_j(\mathbf{p})}{b(\mathbf{p})}$
- Then (2.4) and (2.5) can be expressed as

$$\begin{aligned}
&e(\mathbf{p}, u) = a(\mathbf{p}) + ub(\mathbf{p}) \\
&V(\mathbf{p}, y) = A(\mathbf{p})y + B(\mathbf{p}) \\
&q_j(\mathbf{p}, y) = \alpha_j(\mathbf{p}) + \beta_j(\mathbf{p})y \\
&\text{where, } A(\mathbf{p}) = \frac{1}{b(\mathbf{p})} \qquad B(\mathbf{p}) = -\frac{a(\mathbf{p})}{b(\mathbf{p})} \\
&\text{and, } \alpha_j(\mathbf{p}) = \frac{\partial a(\mathbf{p})}{\partial p_j} - \beta_j(\mathbf{p})a(\mathbf{p}) \qquad \beta_j(\mathbf{p}) = \frac{1}{b(\mathbf{p})}\frac{\partial b(\mathbf{p})}{\partial p_j}
\end{aligned} \tag{2.6}$$

- The budget share equations in this case are given by a weighted average of two terms

$$w_j = \left(\frac{a}{y}\right)\left(\frac{p_j a_j}{a}\right) + \left(1 - \frac{a}{y}\right)\left(\frac{p_j b_j}{b}\right), \tag{2.7}$$

- Implications
  - if $a = y$ (subsistence level is equal to the entire income) the budget share of good j is equal to just $\frac{p_j a_j}{a}$, and if expenditure is much larger than the subsistence level (so $a/y \approx 0$) then the share is given by $\frac{p_j b_j}{b}$
  - In aggregate, the expenditure patterns are a weighted average of value shares appropriate to very rich and very poor consumers
  - Engle curves are still linear but they do not go through the origin anymore
  - although homotheticity implies unitary income elasticities for all commodities, quasi-hometheticity implies elasticities that only tend to unity as total expenditure increases
  - significant generalization/improvement over the previous case, but still restrictive as it unlikely to be true for narrowly defined commodities
  - even for broad commodities such as food, household budget studies tend to give nonlinear Engel curves (we will get to that further below)

- Example: Stone-Geary utility/linear expenditure system (LES) – $u(\mathbf{q}) = \prod_j^J (q_j - \alpha_j)^{\beta_j}$ or equivalently as $u(\mathbf{q}) = \sum_j^J \beta_j \ln(q_j - \alpha_j)$ with $\sum_j^J \beta_j = 1$
  - implied expenditure, indirect utility and demand functions are

$$e(\mathbf{p}, u) = \sum_i^J p_j \alpha_j + u \prod_j^J p_j^{\beta_j}, \qquad V(\mathbf{p}, y) = \frac{y - \sum_j^J p_j \alpha_j}{\prod_j^J p_j^{\beta_j}},$$

and

$$q_j(\mathbf{p}, y) = \alpha_j + \beta_j \frac{y - \sum_j^J p_j \alpha_j}{p_j}$$

  - expenditure on good j is

$$p_j q_j = p_j \alpha_j + \beta_j (y - \sum_j^J p_j \alpha_j)$$

  and is called the **linear expenditure system** (LES) (expenditure is linear in prices and income) which is easy to estimate, and has been very popular in empirical studies for this reason

- Example: Stone-Geary utility/linear expenditure system (LES) – $u(\mathbf{q}) = \prod_j^J (q_j - \alpha_j)^{\beta_j}$ or equivalently as $u(\mathbf{q}) = \sum_j^J \beta_j \ln(q_j - \alpha_j)$ with $\sum_j^J \beta_j = 1$
- expenditure on good j is

$$p_j q_j = p_j \alpha_j + \beta_j (y - \sum_j^J p_j \alpha_j)$$

  and is called the **linear expenditure system** (LES) (expenditure is linear in prices and income) which is easy to estimate, and has been very popular in empirical studies for this reason
    - characterized by the marginal budget share and subsistence level parameters, requiring estimation of $2J$ parameters
    - compare that to the more general case of estimating $J^2 + J$ parameters (own and cross-price elasticities and income/expenditure elasticities), or, if adding up, homogeneity, and symmetry restrictions are imposed, there are $(2J - 1)(J/2 + 1)$ parameters to be estimated
    - nonetheless, if concavity of the expenditure function is allowed, then by construction all cross price elasticities are positive and hence the system cannot be used if some of the products are complements
    - also there is an approximate proportionality between own price and expenditure elasticities

- Aggregate demand data raises the problem as to whether the aggregate demand function is consistent with consumer theory
- Certain conditions are necessary under which we can treat aggregate demand estimations as resulting from the behavior of a single utility maximizing consumer (exact aggregation)
- As you can guess by now, they have to do with quasi-homothecity and Gorman Polar Form

- Suppose there are N consumers (or households) that face the same prices but differ only in the incomes or expenditures on different products so that the demand for good j for the n^{th} individual is of form

$$q_{jn} = g_{jn}(\mathbf{p}, y_n). \tag{2.8}$$

Then the average demand $\bar{q}_j$ – aggregated by adding up *quantities* over all individuals and dividing by N – is given by some function f_j as

$$\bar{q}_j = f_j(\mathbf{p}, y_1, y_2, \ldots y_N) = \frac{1}{N}\sum_n^N g_{jn}(\mathbf{p}, y_n) \tag{2.9}$$

exact aggregation is possible if we can write (2.9) in the form

$$\bar{q}_j = g_j(\mathbf{p}, \bar{y}) \text{ where } \bar{y} = \frac{1}{N}\sum_n^N y_n \tag{2.10}$$

- An implication is that the general function in (2.8) must be linear in y_n, that is, for some function α_{jn} and β_j of $\mathbf{p}$ alone, be of form

$$q_{jn}(\mathbf{p}, y_n) = \alpha_{jn}(\mathbf{p}) + \beta_j(\mathbf{p})y_n \tag{2.11}$$

- Thus, if the aggregate (average) demand is a function of prices and average income, as in (2.10), then the underlying individual demand must be of form given by (2.11)
- But this is the same demand function from quasi-homothetic preferences as in (2.6) with a subscript n for the n^{th} consumer, and α_j and y both vary over consumers, but importantly, β_j does not vary over consumers (i.e, person specific $\alpha(\mathbf{p})$ but identical $\beta(\mathbf{p})$)

- Conversely, if the n^{th} consumer has quasi-homothetic preferences with demand given by (2.11), then the average demand – aggregated via adding up *quantities* over all individuals and dividing by N – is

$$\bar{q}_j = \frac{1}{N}\sum_n^N q_{jn}(\mathbf{p}, y_n)$$
$$= \alpha_j(\mathbf{p}) + \beta_j(\mathbf{p})\bar{y}, \text{ where} \tag{2.12}$$
$$\alpha_j(p) = \frac{1}{N}\sum_n^N \alpha_{jn}(\mathbf{p}), \text{ and } \bar{y} = \frac{1}{N}\sum_n^N y_n.$$

- Thus, for exact linear aggregation, underlying individual demand must be from quasi-homothetic preferences and if the consumer has demand corresponding to quasi-homothetic preferences, then aggregate demand must be of similar form
- (2.11) is necessary and sufficient for (2.10)
- Note that the forms above are arising only due to aggregation requirements, and have nothing to do with requiring aggregate utility maximization

- Suppose now that individuals maximize utility and the individuals demand function is of form (2.11)
- Gorman showed that quasi-homothetic demand of the form above is generated by consumer with the expenditure function given by

$$e_n(\mathbf{p}, u_n) = a_n(\mathbf{p}) + u_n b(\mathbf{p}), \tag{2.13}$$

  i.e., expenditure is of (Gorman) polar form with subscript n in equation (2.6)
- Deaton and Muellbaur show that it is a 'if and only if' condition
- Similarly, the average of the expenditure functions in (2.13) is

$$\bar{e}(\mathbf{p}, u_n) = \bar{a}(\mathbf{p}) + ub(\mathbf{p}), \tag{2.14}$$

  and corresponds to expenditure function for the average demand function in (2.12)
- If individuals maximize utility, and preferences are such that they satisfy the exact aggregation condition, then the average demand function will be consistent with utility maximization
- Moral of the story ... if we want exact aggregation and want to think of the aggregate demand as arising from a utility maximization of a aggregate consumer, then we have to work with quasi-homothetic utility functions

- Aggregation given earlier leads to the linear Engel curves.
- Muellbauer (1975,1976) introduced exact nonlinear aggregation by starting with budget shares rather than with quantities, so that aggregation is over the budget shares of different consumers
- For n consumers, the average budget share of good j is given by

$$\bar{w}_j = \frac{p_j \sum_n q_{jn}(\mathbf{p}, y_n)}{\sum_n y_n} = \sum_n \left(\frac{y_n}{\sum_n y_n} \right) w_{jn}. \qquad (2.15)$$

  defined as a weighted average of individual shares w_{jn} with weights given by the share of each individual in total expenditure on good j.
- Turns out that such a representative consumer (and the assumed cost function) exists only if the preferences are such that the expenditure function of each individual has the form (called **Generalized Gorman Polar Form**)

$$e_n(\mathbf{p}, u_n) = \theta_n(u_n, a(\mathbf{p}), b(\mathbf{p})) + \phi_n(\mathbf{p}) \qquad (2.18)$$

  where $a(\mathbf{p}), b(\mathbf{p})$ and $\phi(\mathbf{p})$ are homogenous of degree 1 in prices, $\theta_n(\ )$ is homogenous in $a(\mathbf{p})$ and $b(\mathbf{p})$ and, $\sum_n \phi_n(\mathbf{p}) = 0$

- Deaton and Muellbauer consider a special case, in which the representative consumers expenditure level (income) y_0 is assumed to depend on the distribution of individual expenditures (incomes), $y_1, \ldots, y_n$ but not on prices, which leads to particularly useful class of demand equations
- For a representative consumer the expenditure function takes the form

$$e(\mathbf{p}, u_0) = [a(\mathrm{p})^{\alpha}(1 - u_0) + b(\mathbf{p})^{\alpha} u_0]^{1/\alpha} \tag{2.24}$$

  and the corresponding budget share equations are said to have the **price independent generalized linear** form (PIGL).
- As $\alpha \to 0$, the representative expenditure function becomes

$$\ln(e(\mathbf{p}, u_0)) = (1 - u_0)\ln(a(\mathbf{p})) + u_0 \ln(b(\mathbf{p})) \tag{2.25}$$

- These give the nonlinear Engel curves as

$$w_j = \begin{cases} \gamma_j + \eta_j (y/k)^{-\alpha} & \text{PIGL} \\ \gamma_j^* + \eta_j^* \ln(y/k) & \text{PIGLOG} \end{cases} \tag{2.26}$$

  where γ's and η's are functions of prices only, k varies over individuals (or households) and can be used to capture demographic effects

- PIGL/PIGLOG family generates exact nonlinear aggregation over individuals or households with nonlinear Engel curves
- Merits of representation of market demand as if they were the outcome of decisions by a rational representative consumer has made for extensive application of this class of models
- A specific application comes from a second-order Taylor series expansion of equation (2.25) so that the first and second derivatives of the expenditure function with respect to prices and utility can be set equal to those of any arbitrary expenditure function at any point (a flexible functional form)
- Deaton and Muellbauer suggest functional forms for $a(\mathbf{p})$ and $b(\mathbf{p})$ in (2.25) which result in a flexible system they call the 'almost ideal demand system', where

$$\begin{aligned} \ln a(p) &= \alpha_0 + \sum_j \alpha_j \ln p_j + \frac{1}{2} \sum_j \sum_k \gamma^*_{jk} \ln p_j \ln p_k \\ \ln b(p) &= \ln a(p) + \beta_0 \prod_j p_j^{\beta_j} \end{aligned} \tag{2.27}$$

- AIDS expenditure function is given by

$$\ln e(\mathbf{p}, u) = \alpha_0 + \sum_j \alpha_j \ln p_j + \frac{1}{2} \sum_j \sum_k \gamma^*_{jk} \ln p_j \ln p_k + u\beta_0 \prod_j p_j^{\beta_j} \tag{2.28}$$

- The expenditure function will be linearly homogenous in $\mathbf{p}$ as long as $\sum_j \alpha_j = 1, \sum_j \gamma^*_{kj} = \sum_k \gamma^*_{kj} = \sum_j \beta_j = 0$

- AIDS demand functions in budget share form are

$$w_j = \alpha_j + \sum_k \gamma_{jk} \ln p_k + \beta_j \ln(y/P)$$

where P is a price index defined by (2.29)

$$lnP = \alpha_0 + \sum_k \alpha_k \ln p_k + \frac{1}{2} \sum_i \sum_k \gamma_{ki} \ln p_k \ln p_i$$

where $\gamma_{jk} = \frac{1}{2}(\gamma^*_{jk} + \gamma^*_{kj})$

- The restrictions on the parameter of the cost function impose restriction on the parameters of the AIDS demand system (2.29) given by

$$\sum_{j=1}^{J} \alpha_j = 1 \qquad \sum_{j=1}^{J} \gamma_{jk} = 0 \qquad \sum_{j=1}^{J} \beta_j = 0$$

$$\sum_k \gamma_{jk} = 0 \qquad \gamma_{jk} = \gamma_{kj} \tag{2.30}$$

- Provided the restrictions above hold (or are imposed), (2.29) represents a system of demand functions which add up to total expenditure ($\sum w_j = 1$), are homogeneous of degree zero in prices and total expenditure taken together, and satisfy Slutsky symmetry and give nonlinear Engle curves.

- Separability refers to the case when a consumer's preferences for products of one group are independent of product specific consumption of products from other groups
- Multi-stage budgeting refers to when a consumer (or household) can allocate their total expenditure on different goods in sequential stages, represented as a utility tree, where in the first stage, the total current expenditure is allocated to broad groups of products (food, housing, entertainment) followed by allocation of expenditures within each broad group (e.g., meats, vegetables, etc. within the food group)

- **Separability** Preferences for products of one group are independent of product specific consumption of products from other groups
- Thus,

$$u(q_1, \ldots, q_j) = f[v_1(\mathbf{q_{(1)}}), \ldots, v_k(\mathbf{q_{(k)}}), \ldots v_K(\mathbf{q_{(K)}})], \quad (2.31)$$

where $(q_1, \ldots, q_j) = (\mathbf{q_{(1)}}, \mathbf{q_{(2)}}, \ldots, \mathbf{q_{(k)}})$ i.e., the set $\{\mathbf{q_{(j)}}\}$ is a partition of $(q_1, \ldots, q_j)$ and there are $K < J$ partitions and $f(\cdot)$ is an increasing function of sub-utility functions $v_1, \ldots, v_k$ defined over the partitions
- The groups could be broad categories such as food, shelter, etc. or within a class of related products it could be sub groups such as type of food (meat, vegetables, etc.)
- This does not remove the dimensionality problem but does lessen it. For example, for a linear demand system, the total number of parameters reduces from $J^2 + J$ (additional J parameters are for income) to $J^2/K + K^2$ number of parameters (for $J = 20$ products and $K = 10$ subgroups, we go from a total of 420 parameters to 140 parameters)

- The implied subgroup demand functions – conditional demand functions – for all products j in group G are of the form

$$q_j = g(y_g, \mathbf{p_g}), \tag{2.32}$$

  where $y_g = \sum_{i \in G} p_i q_i$ is the total expenditure on products in group G and $\mathbf{p_g}$ is the vector of prices of these products
- Note that they do not include the prices of products not in group G
- Let $s_{ij} = \partial q_i^h / \partial p_j$ be the terms of the Slutsky matrix (i.e., partials of the hicksian demand function with respect to prices), then for any two product $i \in G$ and $j \in H$ where $H \neq G$,

$$\begin{aligned} s_{ij} = \mu_{GH} \frac{\partial q_i}{\partial y} \frac{\partial q_j}{\partial y} = \lambda_{GH} \frac{\partial q_i}{\partial y_g} \frac{\partial q_j}{\partial y_h} \\ \text{where } \lambda_{GH} = \mu_{GH} \frac{\partial y_g}{\partial y} \frac{\partial y_h}{\partial y} \end{aligned} \tag{2.33}$$

- μ_{GH} summarizes the interrelation between groups
- λ_{GH} is the compensated derivative of expenditure on group G with respect to a proportional change in all prices in group H (i.e., $\lambda_{GH} = \sum_{j \in H} p_j \frac{\partial y_g}{\partial p_j}\Big|_{u=const}$)
- If there are K total groups, then we can write a $K \times K$ matrix from the $\lambda' s$ that is interpretable as the Slutsky substitution matrix of the group aggregates
- Weak separability results in a two-tier structure of substitution matrices: there are K completely general intragroup Slutsky matrices with no restrictions on substitutions within each group, but between groups substitution is limited by (2.33)

- When the marginal rate of substitution between any two goods belonging to the same group is independent of the consumption of goods within the other groups, it is consider as **weak** separability of preferences
- If the marginal rate of substitution between any two goods belonging to two different groups is independent of the consumption of any good in any third group, this separability is called strong separability or block additivity.[2]
- Strong form is when

$$u(q_1, \ldots, q_j) = f[v_1(\mathbf{q_{(1)}}) + \ldots + v_k(\mathbf{q_{(k)}}) + \ldots + v_K(\mathbf{q_{(K)}})], \tag{2.34}$$

and $f'(\cdot) > 0$. In turn, the equivalent form of (2.33) is given by

$$s_{ij} = \mu \frac{\partial q_i}{\partial x} \frac{\partial q_j}{\partial x} \tag{2.35}$$

where note that μ is independent of groups to which i and j belong

[2] Note that some authors refer to this form as just 'additive' separability (without the use of the word block), but technically that is the case when there is only one good in each group.

- **Multi-stage Budgeting:** Consumers can allocate total expenditures in stages, starting with the top level group and then to any subgroups or sub-subgroups within them
- At each stage, information appropriate for that stage only is required, i.e., the allocation decision is a function of only that group's total expenditure and price indexes for the subgroups and not of prices or price indexes of products in the other groups
  - If the first stage consists of broad categories (food, housing, entertainment) then the consumer decides how much of the budget to allocate to each of these categories depending on three price indexes and not individual prices of types of food items etc
  - Within the food category, the consumer decides how much to spend on different food items (or subgroups) based on the total amount allocated for food and prices of individual food items (or price indexes if there are further subgroups with the food group)
  - Similarly, allocations are done within other groups (housing, entertainment)
  - The process repeats at a third level if there are subgroups (for instance, within foods group, may have subgroups of meat, vegetables, etc., and then within any of these subgroups there are individual items)
- Thus the consumer can allocate the expenditures to the subgroups in sequential stages
- However, all these sequential allocations must equal those that would occur if the consumers utility maximization problem was done in one complete information step

- Because expenditure allocation to any good within a group can be written as a function only of the total group expenditure and the prices of goods within that group, the demand for any good belonging to the group must also be expressed as a function only of total expenditures on the group and the prices of goods within the group
- Thus

$$u(q_1, \ldots, q_m, q_v, q_d, \ldots, q_j) = f[v_1(\mathbf{q_{(1)}}), \ldots, v_F(q_m, q_v, q_d), \ldots, v_K(\mathbf{q_{(K)}})] \tag{2.36}$$

implies

$$q_j = g(y_F, p_m, p_v, p_d) \qquad j \in \{m, v, d\} \tag{2.37}$$

where y_F is the total expenditure on the food items

- Infact, converse is also true: existence of subgroup demand functions implies weak separability

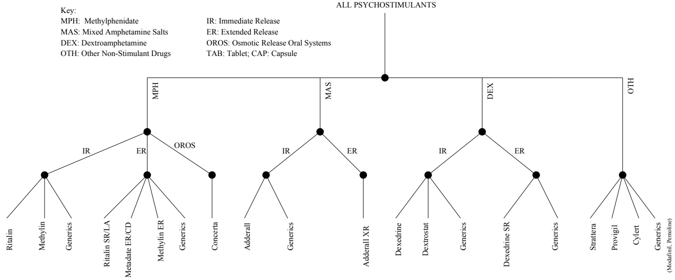
- Weak separability and multi-stage budgeting are closely related concepts but are not the same nor does one imply the other
- Weak separability is necessary and sufficient for the last stage of multi-stage budgeting
- While weak separability is necessary and sufficient for the last stage of multi-stage budgeting, and one can proceed with group specific demand functions at the bottom level (as above)
- Allocation of total budget to different groups at higher stages requires further restrictions on preferences, or on stronger notions of separability and on composite commodity theorem
- To be able to do upper level allocation, there must be an aggregate quantity and price index for each group which can be calculated without knowing the choices within the group
- A useful set of requirements is that (1) the overall utility is separably additive in the sub-utilities, and that (2) the indirect utility functions for each group are of the generalized Gorman polar form

- Typical applications involve a three (or four) stage system where
    - The top level corresponds to overall demand for the product (e.g., beer, pharmaceutical drugs, RTE cereals etc.)
    - The middle level consists of demand for different market segments (e.g., in the demand for beer example, the middle segment consists of four groups of beer – premium beer, light beer, imported beer and non-premium beer, while in the RTE cereal example, the middle segments are family, kids and adults cereals)
    - The bottom level segment involves a flexible brand demand system corresponding to the competition between the different brands within each segment
- For each of these stages a flexible parametric functional form is assumed
- The choice of functional form is driven by the need for flexibility, but also requires that the conditions for multistage budgeting are met

Estimation Details

Setup and Specifications – Tree/Nesting Structure

- We will use a four level system example from Bokhari and Fournier (2013)
  - The top level consists of aggregate demand for drugs used in the treatment of ADHD
  - The second level segments by the types of molecules used in different drugs (four different groups of molecules)
  - The third level further segments the market by the form of the drug, i.e., if it is 4hr, 8hr or a 12hr effect drug
  - The the bottom level, different brands and generics are considered within each molecule-form segment of the market



Note: Generics refer to several manufacturers for each molecule and form given in the column. There are no generic versions of Concerta and Adderall XR during the study period.

- A typical application has the AIDS model at the lowest level
- The demand for product i in segment fm, which consists of I_{fm} number of products, in area a at period t is given by

Level 1 (Bottom):

$$s_{iat_{fm}} = \alpha_{i_{fm}} + \beta_{i_{fm}} ln(\frac{R_{fmat}}{P_{fmat}}) + \sum_{j=1}^{I_{fm}} \gamma_{ij_{fm}} \ln P_{jat_{fm}} + \mathbf{x}_{iat_{fm}} \boldsymbol{\lambda}_{i_{fm}} + \varphi_{iat_{fm}} \tag{3.1}$$

- where
  - $s_{iat_{fm}}$ is the revenue share of product i
  - $\ln P_{jat_{fm}}$ is the (log) price of product j (also in segment f-m)
  - R_{fmat} is the total expenditure on the segment
  - P_{fmat} is a price index for the segment
  - $\mathbf{x}_{iat_{fm}}$ are other exogenous variables which may be varying by product, market or segment and may include terms like demographic variables, time trends, area fixed effects or any observable product characteristics if they vary by markets
- Estimate a system of such equations for each segment, either jointly (all equations from all segments together) or on a segment-by-segment basis – e.g., estimate the system for MPH-IR, MPH-ER, MAS-IR etc.

- The demand for product i in segment fm, which consists of I_{fm} number of products, in area a at period t is given by

Level 1 (Bottom):

$$s_{iat_{fm}} = \alpha_{i_{fm}} + \beta_{i_{fm}} ln(\frac{R_{fmat}}{P_{fmat}}) + \sum_{j=1}^{I_{fm}} \gamma_{ij_{fm}} \ln P_{jat_{fm}} + \mathbf{x}_{iat_{fm}} \boldsymbol{\lambda}_{i_{fm}} + \varphi_{iat_{fm}} \tag{3.1}$$

- To impose the restrictions, we require (for each segment)

$$\sum_{i=1}^{I_{fm}} \alpha_{i_{fm}} = 1 \qquad \sum_{i=1}^{I_{fm}} \gamma_{ik_{fm}} = 0 \qquad \sum_{i=1}^{I_{fm}} \beta_{i_{fm}} = 0$$
$$\sum_{k} \gamma_{ik_{fm}} = 0 \qquad \gamma_{ik_{fm}} = \gamma_{ki_{fm}} \tag{3.2}$$

where the last share equation per segment is not estimated as the shares must add up to one (recall that the revenue shares are shares relative to total spending in this segment and not total spending on all drugs)

- The demand for product i in segment fm, which consists of I_{fm} number of products, in area a at period t is given by

Level 1 (Bottom):

$$s_{iat_{fm}} = \alpha_{i_{fm}} + \beta_{i_{fm}} ln(\frac{R_{fmat}}{P_{fmat}}) + \sum_{j=1}^{I_{fm}} \gamma_{ij_{fm}} \ln P_{jat_{fm}} + \mathbf{x}_{iat_{fm}} \boldsymbol{\lambda}_{i_{fm}} + \varphi_{iat_{fm}} \tag{3.1}$$

- Price Index: Deaton and Muellbaur's exact price index P_{fmat} is given by

$$lnP_{fmat} = \alpha_{0_{fm}} + \sum_{i}^{I_{fm}} \alpha_{i_{fm}} \ln P_{iat_{fm}} + \frac{1}{2} \sum_{i}^{I_{fm}} \sum_{k}^{I_{fm}} \gamma_{ki_{fm}} \ln P_{kat_{fm}} \ln P_{iat_{fm}} \tag{3.3}$$

  This index involves the same parameters that need to be estimated, and hence AIDS estimation requires non-linear estimation methods

University of East Anglia

- The demand for product i in segment fm, which consists of I_{fm} number of products, in area a at period t is given by

Level 1 (Bottom):

$$s_{iat_{fm}} = \alpha_{i_{fm}} + \beta_{i_{fm}} ln(\frac{R_{fmat}}{P_{fmat}}) + \sum_{j=1}^{I_{fm}} \gamma_{ij_{fm}} \ln P_{jat_{fm}} + \mathbf{x}_{iat_{fm}} \boldsymbol{\lambda}_{i_{fm}} + \varphi_{iat_{fm}} \tag{3.1}$$

  - Price Index: Deaton and Muellbaur's exact price index P_{fmat} is given by

$$lnP_{fmat} = \alpha_{0_{fm}} + \sum_{i}^{I_{fm}} \alpha_{i_{fm}} \ln P_{iat_{fm}} + \frac{1}{2} \sum_{i}^{I_{fm}} \sum_{k}^{I_{fm}} \gamma_{ki_{fm}} \ln P_{kat_{fm}} \ln P_{iat_{fm}} \tag{3.3}$$

    This index involves the same parameters that need to be estimated, and hence AIDS estimation requires non-linear estimation methods
  - Alternatively, use **Stone price index**

$$\ln P_{fmat} = \sum_{i}^{I_{fm}} s_{iat_{fm}} \ln P_{iat_{fm}} \tag{3.4}$$

    which makes the estimation linear – but now equation (3.1) involves shares on both

- The demand for product i in segment fm, which consists of I_{fm} number of products, in area a at period t is given by

Level 1 (Bottom):

$$s_{iat_{fm}} = \alpha_{i_{fm}} + \beta_{i_{fm}} ln(\frac{R_{fmat}}{P_{fmat}}) + \sum_{j=1}^{I_{fm}} \gamma_{ij_{fm}} \ln P_{jat_{fm}} + \mathbf{x}_{iat_{fm}} \boldsymbol{\lambda}_{i_{fm}} + \varphi_{iat_{fm}} \tag{3.1}$$

  - Alternatively, use **Stone price index**

$$\ln P_{fmat} = \sum_{i}^{I_{fm}} s_{iat_{fm}} \ln P_{iat_{fm}} \tag{3.4}$$

  which makes the estimation linear – but now equation (3.1) involves shares on both the left hand side and right hand side of the equation

  - In the price index, replace observed shares with average shares
    - In (3.4), Hausman and colleagues replace $s_{iat_{fm}}$ with $\bar{s}_{ia_{fm}}$ – area specific average value of $s_{iat_{fm}}$, thus the value is different for each city but the same for all periods (data is from many periods and a few cities)
    - In (3.4), B&F replace $s_{iat_{fm}}$ with $\bar{s}_{it_{fm}}$ – period specific average value of $s_{iat_{fm}}$, thus the value is different for each period but the same for all areas (data is from many counties and a few periods)

- At the next level up (the middle level, or level 2), demand captures the allocation between segments and can again be modeled using the AIDS specification, in which case the demand specified by equation (3.1) is used with both expenditure shares and prices aggregated to a segment level
- Level 2 is aggregation up from level 1
- Prices are aggregated using either equations (3.3) or (3.4) (exact or Stone price index)
- If the latter (Stone price index) is used, then use $s_{iat_{fm}}$ for the purpose of creating a price index for the upper level rather than $\bar{s}_{at_{fm}}$ or $\bar{s}_{it_{fm}}$

- An alternative for level 2 is lhe log-log equation used by Hausman, Leonard, and Zona (1994) and Hausman (1996) and is given by

Level 2 (Middle):

$$\ln(q_{[fm]at}) = A_{[fm]} + B_{[fm]}\ln(R_{at}) + \sum_{n=1}^{FM}\Gamma_{[fm]n}\ln P_{nat} + \mathbf{x}_{[fm]at}\boldsymbol{\lambda}_{[fm]} + \xi_{[fm]at} \tag{3.5}$$

- where (suppressing subscripts at for areas and periods)
  - $q_{[fm]}$ is the aggregate quantity of the $[fm]$ bottom level segment, i.e., total quantity of RTE cereals for the family, kids or the adults segments in market at (city and quarter)
  - $P_{[fm]}$ is the price of each of these $[fm]$ segments, written as $\ln P_n$ in the equation above, where n is an indexing number for the lower level $[fm]$ segment
  - The segment level prices are the price indexes from the lower level equations and are computed using equations (3.3) or (3.4) as discussed earlier
  - The variable R_{at} is the total expenditure by market on all related products – e.g., it is the sum of total sales of RTE cereals over the the three segments, kids, family and adults
  - And $\mathbf{x}_{[fm]at}$ are the exogenous variables that are segment specific characteristics – if they are different for each market – or just demographic variables by markets

Estimation Details

Setup and Specifications – Middle Level

- Note that lower level of the demand system is AIDS, which satisfies the generalized Gorman polar form,
- In order to be consistent with exact two-stage budgeting, the preferences of the second level should be additively separable (i.e., overall utility from ready-to-eat cereal or all ADHD drugs should be additively separable in the sub-utilities from the various subsegments)
- Neither the second level AIDS, nor the log-log system satisfy this requirement
  - Deaton and Muellbauer also discuss approximate – instead of exact – two-stage budgeting, and show that if one uses the Rotterdam model, approximate two-stage budgeting implies that higher stages also have Rotterdam functional form
- In order for exact multistage budgeting to hold to the next level of aggregation, discussed later, these preferences should be of generalized Gorman polar form

- B&F have two middle level segments which differentiate drugs by forms within molecules (level 2) and by molecules among all ADHD drugs (level 3)

Level 2 (Middle):

$$u_{fat_m} = a_{f_m} + b_{f_m} \ln(\frac{R_{mat}}{P_{mat}}) + \sum_{h=1}^{F_m} g_{fh_m} \ln P_{hat_m} + \mathbf{x}_{fat_m} \boldsymbol{\lambda}_{f_m} + \mu_{fat_m}$$

(3.6)

Level 3 (Middle):

$$\ln(q_{mat}) = A_m + B_m \ln(R_{at}) + \sum_{n=1}^{M} \Gamma_{mn} \ln P_{nat} + \mathbf{x}_{mat} \boldsymbol{\lambda}_m + \xi_{mat}$$

- where (suppressing subscripts at for exposition)
  - u_{f_m} is revenue share of form f within molecule m
  - P_{h_m} is the price of the form (i.e., the price indexes from level 1 segments) given by – $ln(P_{f_m}) = \sum_{j=1}^{I_{f_m}} s_{i_{f_m}} \ln(P_{j_{f_m}})$
  - The terms $\frac{R_m}{P_m}$ are the total expenditures from all forms within molecule m, and a price index for molecule m where the later is computed (using Stone index form) as

$$\ln(P_m) = \sum_{h=1}^{F_m} u_{f_m} \ln(P_{h_m}) \tag{3.7}$$

- For level 2, one needs to estimate as many equations as there are forms per molecule (F_m), and repeat the process for each molecule
  - For instance, if there are four molecules, and each admits up to three forms, then a total of four sets of system equations, with each set consisting of three equations need to be estimated
  - Again, depending on the data, the estimations can be joint for all segments, or segment by segment, and restrictions can be imposed within each segment much like the lower levels
- Level 3 is an aggregation from level 2
  - Thus, $\ln q_m$ is the aggregate quantity for segment m and is the sum of quantities over all forms within this molecule
  - Similarly, $\ln P_n$ is the price of molecule n used earlier in level 2 and is given by (3.7)
  - Total number of equations to be estimated equals the number of upper level segments, e.g., total number of molecules and the rest is the same as discussed earlier in the context of middle level equation (3.5)

- The top level is the demand for the entire set of subsegments (RTE cereal, beer, ADHD drugs etc.) and is typically specified as

Level 4 (Top):

$$\ln Q_{at} = A + B\ln(Y_{at}) + G\ln P_{at} + \mathbf{x}_{at}\boldsymbol{\lambda} + \zeta_{at} \quad (3.8)$$

- where
  - q_{at} is the total quantity
  - Y_{at} is the real income
  - $\mathbf{x}_{at}$ are the demand shifters
  - and P_{at} is the overall price index for these products, given by share weighted sum of (log) prices at the previous level and given by (again suppressing subscripts at),

$$\ln(P) = \sum_{m=1}^{M} v_m \ln(P_m) \quad (3.9)$$

  - and where v_m is the revenue share and P_m is the price index for molecule m computed earlier in (3.7). Note that this form does satisfy additive separability, which is required for exact two-stage budgeting.
- Note that every time we move up one level up, the price index from the lower level is the 'price' at the higher level – and the 'price' at the higher level is constructed as share weighted average (NOT average fixed share)
- Note that this form does satisfy additive separability, which is required for exact two-stage budgeting

- Multi-budgeting process allows estimation of the conditional demand functions (conditional on expenditures on the segment) at the lower levels and the cross-price elasticities are limited to within the segment
- From these conditional demand estimates, and estimates of the upper level equations, it is possible to derive the unconditional cross-price elasticities across the full range of products in different segments
- Conditional on segment expenditure R_{fm} (in market at), price elasticity of a product is

$$\frac{\partial lnQ_{i_{fm}}}{\partial lnP_{k_{f'm'}}} = \frac{1}{s_{i_{fm}}}\Big\{\Big(-\beta_{i_{fm}}\bar{s}_{k_{f'm'}} + \gamma_{ij_{f'm'}}\Big)\cdot 1[f'=f, m'=m]\Big\} - 1[i=k, f'=f, m'=m], \tag{3.10}$$

- where
  - $1[\cdot]$ is the indicator function
  - elasticities conditional on R_{fm} are zero across products in different f-m segments
  - the subscript at has been suppressed in the equation above but is present on all quantities, shares, prices etc. and $\bar{s}_{k_{f'm'}}$ is either $\bar{s}_{kt_{f'm'}}$ or $\bar{s}_{ka_{f'm'}}$ depending on whichever one was used in the Stone price index in level 1 share equations
  - elasticities can be computed in each market or at the average value of shares

- Elasticity at level 2 with respect to the *price index* for the segment and conditional on segment revenue R_m in market at (where the market subscripts have been suppressed), has a similar formula as for the bottom level (since both are in AIDS form) and is given by

$$\frac{\partial lnQ_{fm}}{\partial lnP_{f'_{m'}}} = \frac{1}{u_{fm}}\left\{\left(- b_{fm}\bar{u}_{f'_{m'}} + g_{fh_{m'}}\right) \cdot 1[m' = m]\right\} - 1[f' = f, m' = m], \tag{3.11}$$

- Conditional cross price elasticity of forms in different level 3 segments (i.e., for forms in different molecules) is zero
- Price elasticities at level 3 (for example, at the molecule level), are just the Γ_{mn} parameters in level 3 equation,
- Elasticity with respect to price for the aggregate product is the value of the parameter G in top level equation

- Given all the parameters, unconditional elasticities can be computed as

$$\begin{aligned}\frac{\partial lnQ_{i_{fm}}}{\partial lnP_{k_{f'_{m'}}}} &= \left(1+\frac{\beta_{i_{fm}}}{s_{i_{fm}}}\right)\bar{s}_{k_{f'_{m'}}}\left[\frac{g_{ff'_{m'}}}{u_{fm}}+\bar{u}_{f'_{m'}}\right]\cdot 1[m=m'] \\ &+\left(1+\frac{\beta_{i_{fm}}}{s_{i_{fm}}}\right)\bar{s}_{k_{f'_{m'}}}\left[\frac{b_{fm}\bar{u}_{f'_{m'}}}{u_{fm}}+\bar{u}_{f'_{m'}}\right]\Gamma_{mm'} \\ &+\frac{1}{s_{i_{fm}}}\left\{\gamma_{ik_{f'_{m'}}}-\beta_{i_{fm}}\bar{s}_{k_{f'_{m'}}}\right\}\cdot 1[f'=f,m'=m] \\ &-1[i=k,f'=f,m'=m]\end{aligned} \tag{3.12}$$

- Earlier we discussed how endogeneity can arise in the context of a competitive single product demand-supply model, where due to the simultaneity, the price and the error term in the demand equation are correlated (see equation (1.15))
- The 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

$$\Pi_l = \sum_{r \in \mathfrak{L}_l} (p_r - c_r) q_r(\mathbf{p}, \mathbf{z}_r, \xi_r), \tag{3.13}$$

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

$$q_i(\mathbf{p}, \mathbf{z}_i; \xi_i) + \sum_{r \in \mathfrak{L}_l} (p_r - c_r) \frac{\partial q_r(\mathbf{p}, \mathbf{z}_r; \xi_r)}{\partial p_i} = 0 \tag{3.14}$$

- 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

$$\mathbf{p} = \mathbf{c} + \boldsymbol{\Omega}^{-1} \boldsymbol{q}(\boldsymbol{p}, \boldsymbol{z}; \boldsymbol{\xi}), \tag{3.15}$$

- 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 **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

- A second set of instruments is due to Hausman et al. (1994) and has been used in several papers – **common cost shocks** and **no common demand side shocks** across cities
- The reduced form price of a product i in two cities, $a = 1$ and $a = 2$ at time period t, will be given by

$$\begin{aligned} \ln p_{i1t} &= \pi_1 \ln c_{it} + \mathbf{x}_{i1t}\boldsymbol{\pi}_2 + v_{i1t} \\ \ln p_{i2t} &= \pi_1 \ln c_{it} + \mathbf{x}_{i2t}\boldsymbol{\pi}_2 + v_{i2t}, \end{aligned} \tag{3.16}$$

- where
  - c_{it} is the common cost component of the price in two different cities
  - $\mathbf{x}_{iat}$ are brand level demand shifters (demographics, time trends) as well city specific brand differentials (intercepts by brands and cities) due to differences in transportation costs or local wages
  - In general, the error terms v_{iat} will be correlated with φ_{iat} in the equation (3.1), and hence OLS/SUR will give inconsistent estimates
  - If however, v_{i1t} is uncorrelated with v_{i2t}, then city two's prices will be uncorrelated with the error term φ_{i1t} in the equation (3.1), and hence the instrument will be valid
  - Further, since the prices in the two cities are driven by the same underlying common costs c_{it}, they will be correlated to each other and hence relevant
- Hausman instruments also rely on no correlation between v_{i1t} and v_{i2t} – this assumption may be in invalid if the 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

Characteristic Space
General Idea

- 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

Characteristic Space
General Idea

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

$$u_{njt} = U(\mathbf{x}_{jt}, \xi_{jt}, y_{nt} - p_{jt}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{njt}; \boldsymbol{\theta}_n), \qquad \text{for } j = 0, 1, 2, \ldots, J \tag{4.1}$$

  - ‘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

$$u_{njt} = U(\mathbf{x}_{jt}, \xi_{jt}, y_{nt} - p_{jt}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{njt}; \boldsymbol{\theta}_n), \qquad \text{for } j = 0, 1, 2, \ldots, J \qquad (4.1)$$

  - 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

$$u_{njt} = U(\mathbf{x}_{jt}, \xi_{jt}, y_{nt} - p_{jt}, \boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{njt}; \boldsymbol{\theta}_n), \qquad \text{for } j = 0, 1, 2, \ldots, J \tag{4.1}$$

  - 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

$$\mathbb{A}_{jt}(\boldsymbol{\theta}) = \{(\boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{n0t}, \epsilon_{n1t}, \ldots, \epsilon_{nJt}) \mid u_{njt} > u_{nlt} \qquad \forall\, l = 0, 1, 2 \ldots J,\ l \neq j\} \tag{4.2}$$

- 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 $\mathbb{A}_{jt}$,

$$s_{jt}(\mathbf{x}, \mathbf{p}; \boldsymbol{\theta}) = \int_{\mathbb{A}_{jt}} dF(\boldsymbol{d}, \boldsymbol{\nu}, \boldsymbol{\epsilon}). \tag{4.3}$$

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

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

$$
\begin{aligned}
&u_{njt} = \alpha_n(y_n - p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta}_n + \xi_{jt} + \epsilon_{njt}, \text{ where} \\
&n = 1, \ldots, N, \qquad j = 0, 1, \ldots, J, \qquad t = 1, 2, \ldots, T, \text{ and} \\
&\boldsymbol{\beta}_n = \boldsymbol{\beta}, \qquad \alpha_n = \alpha, \qquad \text{for all } N
\end{aligned}
\tag{5.1}
$$

- 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

$$
u_{n0t} = \alpha y_n + \epsilon_{n0t} \tag{5.2}
$$

- Utility function in (5.1) can be written more compactly as just

$$u_{njt} = \alpha y_n + \delta_{jt} + \epsilon_{njt}, \tag{5.3}$$

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

$$\mathbb{A}_{jt}(\alpha, \beta) = \{(\epsilon_{n0t}, \epsilon_{n1t}, \ldots \epsilon_{nJt}) | u_{njt} > u_{nlt} \qquad \forall\, l = 0, 1, 2 \ldots J,\; l \neq j\} \tag{5.4}$$

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

$$f(\epsilon) = \exp(-\epsilon)\exp(-\exp(-\epsilon)) \text{ and } \qquad F(\epsilon) = \exp(-\exp(-\epsilon)), \tag{5.5}$$

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

$$s_{jt}(\boldsymbol{\delta}_t) = \int_{\mathbb{A}_{jt}} dF(\boldsymbol{\epsilon}) = \frac{\exp(\delta_{jt})}{\sum_{j=0}^{J} \exp(\delta_{jt})}. \tag{5.6}$$

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

$$\begin{aligned} s_{jt} &= \frac{\exp(\delta_{jt})}{1 + \sum_{j=1}^{J} \exp(\delta_{jt})} \\ s_{0t} &= \frac{1}{1 + \sum_{j=1}^{J} \exp(\delta_{jt})} = 1 - \sum_{j=1}^{J} s_{jt}. \end{aligned} \tag{5.7}$$

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

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \tag{5.8}$$

can be estimated using linear regression methods

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

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \tag{5.8}$$

  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 (5.6))

$$\Pr(j) = \frac{\exp(\delta_j)}{\sum_{j=0}^{J} \exp(\delta_j)} \tag{5.9}$$

The relative probabilities of options j and k are thus

$$\frac{\Pr(j)}{\Pr(k)} = \frac{\exp(\delta_j)}{\exp(\delta_k)} = \exp(\delta_j - \delta_k) \tag{5.10}$$

- 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

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k, \\ \alpha p_{kt} s_{kt} & \text{otherwise} \end{cases} \tag{5.11}$$

- 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

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k, \\ \alpha p_{kt} s_{kt} & \text{otherwise} \end{cases} \tag{5.11}$$

- 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 (5.8) 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 $\times$ 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

$$s_{jt} = q_{jt}/M_t \qquad j = 1, \ldots, J \text{ for all } t = 1, \ldots, T. \tag{5.12}$$

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

$$s_{0t} = 1 - \sum_{j=1}^{J} s_{jt} \qquad \forall t \tag{5.13}$$

- With these definitions in place, can estimate equation (5.8) (reproduced below)

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} \equiv \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}, \tag{5.8}$$

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

- With these definitions in place, can estimate equation (5.8) (reproduced below)

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} \equiv \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}, \tag{5.8}$$

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

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

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}_t = \begin{bmatrix} \ln s_1 - \ln s_0 \\ \ln s_2 - \ln s_0 \\ \vdots \\ \ln s_J - \ln s_0 \end{bmatrix}_t = \alpha \begin{bmatrix} -p_1 \\ -p_2 \\ \vdots \\ -p_J \end{bmatrix}_t + \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1K} \\ x_{21} & x_{22} & \ldots & x_{2K} \\ \vdots & & & \\ x_{J1} & x_{J2} & \ldots & x_{JK} \end{bmatrix}_t \begin{bmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_J \end{bmatrix} \tag{5.14}$$

- Data from multiple markets can be vertically 'stacked'

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

$$\begin{bmatrix} \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{J1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Jt} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} y_{1T} \\ y_{2T} \\ \vdots \\ y_{JT} \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} \ln s_{11} - \ln s_{01} \\ \ln s_{21} - \ln s_{01} \\ \vdots \\ \ln s_{J1} - \ln s_{01} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \ln s_{1t} - \ln s_{0t} \\ \ln s_{2t} - \ln s_{0t} \\ \vdots \\ \ln s_{Jt} - \ln s_{01} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \ln s_{1T} - \ln s_{0T} \\ \ln s_{2T} - \ln s_{0T} \\ \vdots \\ \ln s_{JT} - \ln s_{0T} \end{pmatrix} \end{bmatrix} = \alpha \begin{bmatrix} \begin{pmatrix} -p_{11} \\ -p_{21} \\ \vdots \\ -p_{J1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} -p_{1t} \\ -p_{2t} \\ \vdots \\ -p_{Jt} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} -p_{1T} \\ -p_{2T} \\ \vdots \\ -p_{JT} \end{pmatrix} \end{bmatrix} + \begin{bmatrix} \begin{pmatrix} x_{111} & x_{121} & \dots & x_{1K1} \\ x_{211} & x_{221} & \dots & x_{2K1} \\ \vdots & & & \\ x_{J11} & x_{J21} & \dots & x_{JK1} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} x_{11t} & x_{12t} & \dots & x_{1Kt} \\ x_{21t} & x_{22t} & \dots & x_{2Kt} \\ \vdots & & & \\ x_{J1t} & x_{J2t} & \dots & x_{JKt} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} x_{11T} & x_{12T} & \dots & x_{1KT} \\ x_{21T} & x_{22T} & \dots & x_{2KT} \\ \vdots & & & \\ x_{J1T} & x_{J2T} & \dots & x_{JKT} \end{pmatrix} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} + \tag{5.15}$$

- As discussed in earlier, very likely that $\text{cov}(p_{jt}, \xi_{jt}) \neq 0$
    - As before, one needs to find instruments that are correlated with price but not with any of the unobserved product characteristics
    - See earlier discussion on various instruments (Hausman, BLP etc.)
- 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

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_j + \xi_t + \Delta\xi_{jt} \tag{5.16}$$

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

$$\mathrm{E}(\Delta\xi_{jt}p_{jt}|\mathbf{x}_{jt}) = 0 \tag{5.17}$$

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

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_j + \xi_t + \Delta\xi_{jt} \tag{5.16}$$

- 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 (5.17) is not true

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

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_j + \xi_t + \Delta\xi_{jt} \tag{5.16}$$

- 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

- 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 (5.16)

- 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 (5.16)
  - Then, the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ in equation

$$\mathbf{b}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\xi}_t, \tag{5.18}$$

can be recovered via GLS estimator

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}_t'\hat{\mathbf{V}}_{\mathbf{b}}^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\hat{\mathbf{V}}_{\mathbf{b}}^{-1}\hat{\mathbf{b}}_t, \text{ and } \boldsymbol{\xi}_t = \hat{\mathbf{b}}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}} \tag{5.19}$$

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

$$u_{njt} = \delta_{jt} + \zeta_{ngt}(\sigma) + (1 - \sigma)\epsilon_{njt}, \tag{5.20}$$

- 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

$$\ln(s_{jt}) - \ln(s_{0t}) = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \sigma \ln(s_{jt}/s_{gt}) + \xi_{jt} \tag{5.21}$$

- 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

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + u_t \tag{5.22}$$

- 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

$$\mathrm{E}[\mathbf{x}_t'(y_t - \mathbf{x}_t \boldsymbol{\beta})] = \mathbf{0} \tag{5.23}$$

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

$$\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t'(y_t - \mathbf{x}_t \boldsymbol{\beta}) = \mathbf{0} \tag{5.24}$$

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

$$\widehat{\boldsymbol{\beta}}_{MM} = (\sum_t \mathbf{x}_t'\mathbf{x}_t)^{-1} \sum_t \mathbf{x}_t' y_t = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{5.25}$$

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

$$\mathrm{E}[\mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})] = \mathbf{0} \tag{5.26}$$

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

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \mathbf{0} \tag{5.27}$$

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

$$\widehat{\boldsymbol{\beta}}_{MM} = (\sum_t \mathbf{z}_t'\mathbf{x}_t)^{-1}\sum_t \mathbf{z}_t' y_t = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} \tag{5.28}$$

- 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

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}\sum_t \mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})\right]' \boldsymbol{\Phi} \left[\frac{1}{T}\sum_t \mathbf{z}_t'(y_t - \mathbf{x}_t\boldsymbol{\beta})\right] \tag{5.29}$$

  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 (5.29) becomes

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{Z}\right] \boldsymbol{\Phi} \left[\frac{1}{T}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \tag{5.30}$$

  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

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\left[\frac{1}{T}\mathbf{X}'\mathbf{Z}\right]\boldsymbol{\Phi}\left[\frac{1}{T}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right] = \mathbf{0} \tag{5.31}$$

- Then the GMM linear IV estimator and its variance are

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\text{GMM}} &= \left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{y} \\ \mathrm{V}(\widehat{\boldsymbol{\beta}})_{\text{GMM}} &= T\left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\widehat{\mathbf{S}}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\mathbf{X}\right)^{-1} \end{aligned} \tag{5.32}$$

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

$$\mathbf{S} = \text{plim}\frac{1}{T}\sum_i\sum_j\left[\mathbf{z}_i' u_i u_j \mathbf{z}_j\right] \tag{5.33}$$

- 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:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{GMM}} &= \left(\mathbf{X}'\mathbf{Z}\mathbf{\Phi}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\mathbf{\Phi}\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{\Phi}^{-1}(\mathbf{X}'\mathbf{Z})^{-1}(\mathbf{X}'\mathbf{Z})\mathbf{\Phi}\mathbf{Z}'\mathbf{y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} = \widehat{\boldsymbol{\beta}}_{\text{IV}}\end{aligned} \tag{5.34}$$

- 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

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{GMM}} &= \left(\mathbf{X}'\mathbf{P_z}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{P_z}\mathbf{y}\right) = \widehat{\boldsymbol{\beta}}_{\text{2SLS}} \\ \mathrm{V}(\widehat{\boldsymbol{\beta}}_{\text{GMM}}) &= s^2\left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1} \\ \text{where } \mathbf{P_z} &= \mathbf{Z}(\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z}' \text{ and } s^2 = (T-K)^{-1}\sum_t \hat{u}_t^2\end{aligned} \tag{5.35}$$

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

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

$$\text{and } \widehat{\mathbf{S}} = T^{-1}\sum_t \hat{u}_t^2\mathbf{z}_t\mathbf{z}'_t. \tag{5.36}$$

- 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 $\boldsymbol{\Phi} = \widehat{\mathbf{S}}^{-1}$

$$\widehat{\boldsymbol{\beta}}_{\text{OGMM}} = \left(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y} \tag{5.37}$$

  - 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 (5.37) to compute the estimator
- Variance is given by

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\text{OGMM}}) = T\left(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1} \tag{5.38}$$

- 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

$$\mathrm{E}\left[\mathbf{h}(\mathbf{w}_t, \boldsymbol{\theta}_0)\right] = \mathbf{0} \tag{5.39}$$

  - 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 (5.29)) is

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}\sum_t \mathbf{h}(\mathbf{w}_t, \boldsymbol{\theta})\right]' \boldsymbol{\Phi} \left[\frac{1}{T}\sum_t \mathbf{h}(\mathbf{w}_t, \boldsymbol{\theta})\right] \tag{5.40}$$

  and the corresponding first order conditions are

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left[\frac{1}{T}\sum_t^T \frac{\partial \mathbf{h}_t(\widehat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}}\right] \boldsymbol{\Phi} \left[\frac{1}{T}\sum_t^T \mathbf{h}_t(\widehat{\boldsymbol{\theta}})\right] = \mathbf{0} \tag{5.41}$$

  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

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{T}\mathbf{u}'\mathbf{Z}\right] \boldsymbol{\Phi} \left[\frac{1}{T}\mathbf{Z}'\mathbf{u}\right] \tag{5.42}$$

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

Back to Logits

GMM Estimation

- 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 <u>observed</u> 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 (5.7) 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

$$\tilde{s}_{jt} = \frac{\exp(\delta_{jt})}{1 + \sum_{j=1}^{J} \exp(\delta_{jt})} \tag{5.7}$$

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

$$\min_{\boldsymbol{\theta}_1} \sum_{j=1}^{J} \left[s_{jt} - \tilde{s}_{jt}(\alpha, \boldsymbol{\beta}, \xi_{1t}, \xi_{2t}, \ldots, \xi_{Jt}) \right]^2 \tag{5.43}$$

- 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

- 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

$$\frac{1}{J}\sum_{j} z_{jt}^{(m)}\xi_{jt} = \frac{1}{J}\sum_{j} z_{jt}^{(m)}(\delta_{jt} - \mathbf{x}_{jt}\boldsymbol{\beta} + \alpha p_{jt}) \tag{5.44}$$

- 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 (5.7) – in J unknowns

$$\begin{aligned} s_{1t} &= \tilde{s}_{1t}(\delta_{1t}, \ldots, \delta_{Jt}) \\ s_{2t} &= \tilde{s}_{2t}(\delta_{1t}, \ldots, \delta_{Jt}) \\ &\vdots \\ s_{Jt} &= \tilde{s}_{jt}(\delta_{1t}, \ldots, \delta_{Jt}) \end{aligned} \tag{5.45}$$

(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 (5.44).

(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,

$$\begin{aligned} \delta_{1t} &= \alpha(-p_{1t}) + \mathbf{x}_{1t}\boldsymbol{\beta} + \xi_{1t} \\ \delta_{2t} &= \alpha(-p_{2t}) + \mathbf{x}_{2t}\boldsymbol{\beta} + \xi_{2t} \\ &\vdots \\ \delta_{Jt} &= \alpha(-p_{Jt}) + \mathbf{x}_{Jt}\boldsymbol{\beta} + \xi_{Jt} \end{aligned} \tag{5.46}$$

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

$$\frac{1}{J}\sum_j z_{jt}^{(m)}\xi_{jt} = \frac{1}{J}\sum_j z_{jt}^{(m)}(\widehat{\delta}_{jt} - \mathbf{x}_{jt}\boldsymbol{\beta} + \alpha p_{jt}) \tag{5.47}$$

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 (5.7) 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 (5.45) will be of the form

$$s_{jt} = \tilde{s}_{jt}(\delta_{1t}, \ldots, \delta_{Jt}, \boldsymbol{\theta}_2) \tag{5.48}$$

- 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

$$\boldsymbol{\delta}_t^{h+1} = \boldsymbol{\delta}_t^h + \left[\ln(\mathbf{s}_t) - \ln(\tilde{\mathbf{s}}_t(\boldsymbol{\delta}_t^h; \boldsymbol{\theta}_2))\right] \tag{5.49}$$

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, $\min_{\boldsymbol{\theta}_1} \boldsymbol{\xi}(\boldsymbol{\theta}_1)'\mathbf{Z}\boldsymbol{\Phi}\mathbf{Z}'\boldsymbol{\xi}(\boldsymbol{\theta}_1)$ 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 (5.37) 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

- Let the utility be given by

$$u_{njt} = \alpha_n(y_n - p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta}_n + \xi_{jt} + \epsilon_{njt}, \text{ where}$$
$$n = 1, \ldots, N, \qquad j = 0 \ldots, J, \qquad t = 1 \ldots, T \tag{6.1}$$

- where

$$\begin{bmatrix} \alpha_n \\ \boldsymbol{\beta}_n \end{bmatrix} = \underbrace{\begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix}}_{\boldsymbol{\theta}_1} + \underbrace{\boldsymbol{\Pi d}_n + \boldsymbol{\Sigma \nu}_n}_{\boldsymbol{\theta}_2 = \{\boldsymbol{\Pi}, \boldsymbol{\Sigma}\}}$$
$$= \begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Pi}_\alpha \\ \boldsymbol{\Pi}_{\boldsymbol{\beta}} \end{bmatrix} \boldsymbol{d}_n + \begin{bmatrix} \boldsymbol{\Sigma}_\alpha \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\nu}_{n\alpha} & \boldsymbol{\nu}_{\boldsymbol{n\beta}} \end{bmatrix} \tag{6.2}$$

- and where

$$\boldsymbol{d}_n \sim F_{\boldsymbol{d}}(\boldsymbol{d}) \qquad \boldsymbol{\nu}_n \sim F_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \tag{6.3}$$

  - 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 $\boldsymbol{\Pi d}_n + \boldsymbol{\Sigma \nu}_n$
  - 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 $\{\boldsymbol{\Pi}$ 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

$$\begin{aligned} \alpha_n = \alpha \quad & +\pi_{11}d_{1n} + \pi_{12}d_{2n} + \ldots + \pi_{15}d_{5n} \\ & +\sigma_{11}v_{1n} + \sigma_{12}v_{2n} + \ldots + \sigma_{14}v_{4n} \\ \beta_{kn} = \beta_k \quad & +\pi_{k1}d_{1n} + \pi_{k2}d_{2n} + \ldots + \pi_{k5}d_{5n} \\ & +\sigma_{k1}v_{1n} + \sigma_{k2}v_{2n} + \ldots + \sigma_{k4}v_{4n} \end{aligned} \tag{6.4}$$

- 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.,

$$\underbrace{\begin{bmatrix}\alpha_n \\ \boldsymbol{\beta}_n\end{bmatrix}}_{k\times 1} = \underbrace{\begin{bmatrix}\alpha \\ \boldsymbol{\beta}\end{bmatrix}}_{k\times 1} + \underbrace{\mathbf{\Pi}\boldsymbol{d}_n}_{k\times D \text{ by } D\times 1} + \underbrace{\mathbf{\Sigma}\boldsymbol{\nu}_n}_{k\times k \text{ by } k\times 1} \tag{6.5}$$

  - 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 (6.2) back into (6.1) 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)

$$
\begin{aligned}
&u_{njt} = \alpha_n y_n + \delta_{jt} + \mu_{njt} + \epsilon_{njt} \\
&\text{where,} \\
&\delta_{jt} = \delta(\mathbf{x}_{jt}, p_{jt}, \xi_{jt}; \boldsymbol{\theta}_1) = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \\
&\mu_{njt} = \mu(\mathbf{x}_{jt}, p_{jt}, \boldsymbol{d}_n, \boldsymbol{\nu}_n; \boldsymbol{\theta}_2) = (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)
\end{aligned}
\tag{6.6}
$$

- 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

$$\begin{aligned}
&u_{njt} = \alpha_n y_n + \delta_{jt} + \mu_{njt} + \epsilon_{njt} \\
&\text{where,} \\
&\delta_{jt} = \delta(\mathbf{x}_{jt}, p_{jt}, \xi_{jt}; \boldsymbol{\theta}_1) = \alpha(-p_{jt}) + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} \\
&\mu_{njt} = \mu(\mathbf{x}_{jt}, p_{jt}, \boldsymbol{d}_n, \boldsymbol{\nu}_n; \boldsymbol{\theta}_2) = (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\boldsymbol{d}_n + \boldsymbol{\Sigma}\boldsymbol{\nu}_n)
\end{aligned} \tag{6.6}$$

- 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 (6.6) – $u_{njt} = \alpha_n y_n + \delta_{jt} + \mu_{njt} + \epsilon_{njt}$)
    - 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

$$\mathbb{A}_{jt}(\boldsymbol{\theta}_2) = \{(\boldsymbol{d}_{nt}, \boldsymbol{\nu}_{nt}, \epsilon_{n0t}, \epsilon_{n1t}, \ldots, \epsilon_{nJt}) \,|\, u_{njt} > u_{nlt} \qquad \forall\, l = 0, 1, 2 \ldots J,\ l \neq J\} \tag{6.7}$$

- 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},

$$s_{jt} = \int_{\mathbb{A}_{jt}} dF(\boldsymbol{d}, \boldsymbol{\nu}, \boldsymbol{\epsilon}) = \int_{\mathbb{A}_{jt}} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) dF_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) \tag{6.8}$$

    - 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

$$s_{njt} = \frac{\exp(\delta_{jt} + \tilde{\mu}_{njt})}{\sum_{j=0}^{J} \exp(\delta_{jt} + \tilde{\mu}_{njt})} \tag{6.9}$$

- 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

$$\begin{aligned} s_{jt} &= \int_{\mathbb{A}_{jt}} s_{njt} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \\ &= \int_{\mathbb{A}_{jt}} \left\{ \frac{\exp(\delta_{jt} + \mu_{njt})}{\sum_{j=0}^{J} \exp(\delta_{jt} + \mu_{njt})} \right\} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \end{aligned} \tag{6.10}$$

University of East Anglia

- Price elasticities of market shares are given by

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int_{\mathbb{A}_{jt}} \alpha_n s_{njt}(1 - s_{njt}) dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) & \text{if } j = k, \\ \frac{p_{kt}}{s_{jt}} \int_{\mathbb{A}_{jt}} \alpha_n s_{njt} s_{nkt} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) & \text{otherwise} \end{cases} \tag{6.11}$$

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

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k, \\ \alpha p_{kt} s_{kt} & \text{otherwise} \end{cases} \tag{5.11}$$

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

University of East Anglia

- Let x be some arbitrary random variable[3] 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., $\mathrm{E}[x] = \int x \cdot dF(x)$
  - 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
  - $\mathrm{E}[g(x)] = \int g(x) \cdot dF(x)$
  - 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

[3]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 (6.10), 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

University of East Anglia

- 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,

$$\begin{aligned} \tilde{s}_{jt} &= \int_{A_{jt}} s_{njt} dF_{\boldsymbol{d}}(\boldsymbol{d}) dF_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \\ &= \frac{1}{N_s} \sum_{n}^{N_s} s_{njt} = \frac{1}{N_s} \sum_{n}^{N_s} \left\{ \frac{\exp(\delta_{jt} + \mu_{njt})}{\sum_{j=0}^{J} \exp(\delta_{jt} + \mu_{njt})} \right\} \end{aligned} \tag{6.12}$$

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 $\boldsymbol{\Pi}\boldsymbol{d}_n$ term

$$\begin{bmatrix}\alpha_n \\ \boldsymbol{\beta}_n\end{bmatrix} = \begin{bmatrix}\alpha \\ \boldsymbol{\beta}\end{bmatrix} + \boldsymbol{\Sigma}\boldsymbol{\nu}_n = \begin{bmatrix}\alpha \\ \boldsymbol{\beta}\end{bmatrix} + \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22}\end{bmatrix}\begin{bmatrix}\nu_{1n} \\ \nu_{2n}\end{bmatrix} \tag{6.13}$$

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

$$\begin{aligned}
\alpha_n &= \alpha + \sigma_{11}\nu_{1n} + \sigma_{12}\nu_{2n} \\
\boldsymbol{\beta}_n &= \boldsymbol{\beta} + \sigma_{21}\nu_{1n} + \sigma_{22}\nu_{2n} \\
\mathrm{E}[\alpha_n] &= \alpha \qquad \mathrm{E}[\boldsymbol{\beta}_n] = \boldsymbol{\beta} \\
\mathrm{Var}[\alpha_n] &= \sigma_{11}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{11}\sigma_{12}\mathrm{Cov}[\nu_{1n},\nu_{2n}] + \sigma_{12}^2\mathrm{Var}[\nu_{2n}] \\
\mathrm{Var}[\boldsymbol{\beta}_n] &= \sigma_{21}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{21}\sigma_{22}\mathrm{Cov}[\nu_{1n},\nu_{2n}] + \sigma_{22}^2\mathrm{Var}[\nu_{2n}]
\end{aligned} \tag{6.14}$$

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

$$\begin{aligned}
\alpha_n &= \alpha + \sigma_{11}\nu_{1n} + \sigma_{12}\nu_{2n} \\
\boldsymbol{\beta}_n &= \boldsymbol{\beta} + \sigma_{21}\nu_{1n} + \sigma_{22}\nu_{2n} \\
\mathrm{E}[\alpha_n] &= \alpha \qquad \mathrm{E}[\boldsymbol{\beta}_n] = \boldsymbol{\beta} \\
\mathrm{Var}[\alpha_n] &= \sigma_{11}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{11}\sigma_{12}\mathrm{Cov}[\nu_{1n}, \nu_{2n}] + \sigma_{12}^2\mathrm{Var}[\nu_{2n}] \\
\mathrm{Var}[\boldsymbol{\beta_n}] &= \sigma_{21}^2\mathrm{Var}[\nu_{1n}] + 2\sigma_{21}\sigma_{22}\mathrm{Cov}[\nu_{1n}, \nu_{2n}] + \sigma_{22}^2\mathrm{Var}[\nu_{2n}]
\end{aligned} \tag{6.14}$$

- 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

$$\begin{aligned}\text{Cov}(\alpha_n, \boldsymbol{\beta}_n) &= \text{E}(\alpha_n \boldsymbol{\beta}_n) - \alpha\boldsymbol{\beta} \\ &= \sigma_{11}\sigma_{21}\text{Var}(\nu_{1n}) + \sigma_{12}\sigma_{22}\text{Var}(\nu_{2n}) \\ &+ \sigma_{11}\sigma_{22}\text{Cov}(\nu_{1n}, \nu_{2n}) + \sigma_{12}\sigma_{21}\text{Cov}(\nu_{1n}, \nu_{2n}) \\ &= \sigma_{11}\sigma_{22}\text{Cov}(\nu_{1n}, \nu_{2n}).\end{aligned} \tag{6.15}$$

  - 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 (6.14), 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 $\mathbf{\Pi}$ are also zero – as in the foregoing discussion – this implies that covariances between marginal utilities will also be zero
- However, if terms of $\mathbf{\Pi}$ are not all zero, they will still invoke correlations between the marginal utilities of different characteristics
  - as equation (6.4), reproduced below for this special case of two characteristics and five demographics, shows

$$\begin{aligned} \alpha_n = \alpha \quad & +\pi_{11} d_{1n} + \pi_{12} d_{2n} + \ldots + \pi_{15} d_{5n} \\ & +\sigma_{11}\nu_{1n} + \sigma_{12}\nu_{2n} \\ \boldsymbol{\beta}_n = \boldsymbol{\beta} \quad & +\pi_{21} d_{1n} + \pi_{22} d_{2n} + \ldots + \pi_{25} d_{5n} \\ & +\sigma_{21}\nu_{1n} + \sigma_{22}\nu_{2n} \end{aligned} \tag{6.4}$$

  - 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 $\mathbf{\Pi}$ to zero (at least not all terms) then setting the off diagonals of $\mathbf{\Sigma}$ 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 (6.10), 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

(-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 (6.12) 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
$\Big(\boldsymbol{\xi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'\mathbf{Z}\Big)\boldsymbol{\Phi}\Big(\mathbf{Z}'\boldsymbol{\xi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\Big)$,
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 <u>or brand dummies</u> 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, pages 4171–4276. 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. 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.

Chaudhuri, S., Goldberg, P. K., and Jia, P. (2006). Estimating the effects of global patent protection in pharmaceuticals: a case study of quinolones in India. *American Economic Review*, 96(5):1477–1514.

Deaton, A. and Muellbauer, J. (1980a). An almost ideal demand system. *American Economic Review*, 70(3):312–326.

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

Ellison, S. F., Cockburn, I., Griliches, Z., and Hausman, J. A. (1997). Characteristics of demand for pharmaceutical products: an examination of four cephalosporins. *RAND Journal of Economics*, 28(3):426–446.

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: the case of the U.S. automobile industry. *Econometrica*, 63(4):pp. 891–951.

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.

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

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729.

References

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.