

ECON8853 - Industrial Organization I

Lecture Notes from Julie H. Mortimer's lectures

Paul Anthony Sarkis
Boston College

Contents

0.1	Monopoly and Imperfect competition with homogeneous and differentiated products	5
0.2	Oligopoly with homogeneous products	5
0.3	Differentiated products	5
1	Introduction to Empirical Industrial Organization	6
1.1	Generic Static Analysis	6
1.1.1	Aside on Dynamic Effects	7
1.1.2	Aside on Theoretical IO and Econometrics	7
1.2	Review of the Cournot model	8
1.3	Review of the Bertrand model	9
1.4	Comments on equilibrium notions	9
1.5	Strategic complementarity of diff. products	9
2	Introduction to Demand Estimation	10
2.1	An early example: Bresnahan (1987)	12

2.2	Single product demand	13
2.2.1	Representative agent	13
2.2.2	Heterogeneous agents	13
2.3	Multi-product demand	14
2.3.1	Product Space approach	14
2.3.2	Issues with AIDS	15
2.3.3	Examples: Hausman, Leonard and Zona (1994)	16
2.3.4	When should you choose product space approach?	17
2.4	Multi-product demand: Characteristic Space	17
2.4.1	History of the characteristics approach	18
2.4.2	Formal base model	18
2.4.3	Types of models	19
2.4.4	Data and issues with estimation	21
3	Estimating demand for vertically differentiated products	22
4	Logit, Nested Logit and Multinomial Probit	23
4.1	”Plain-vanilla” Logit	23
4.1.1	Overview	23
4.1.2	Logit and taste variation	25
4.1.3	Too-many-parameters problem	26
4.1.4	Derivatives and Elasticities	27

4.1.5	Independence of Irrelevant Alternatives (IIA)	28
4.1.6	Consumer Surplus	29
4.2	Nested Logit	30
4.2.1	IIA and substitution patterns	31
4.2.2	Research considerations	32
4.2.3	Extensions of the GEV	32
4.3	Multinomial Probit	32
4.4	GMM Estimation Strategy for Product-level data	33
4.4.1	Inversion in the logit case	33
4.4.2	Inversion in the nested-logit case	34
4.4.3	Endogeneity issues	34
4.4.4	Measurement errors	35
4.4.5	Adding supply-side restrictions	35
4.5	Individual-level data	35
5	Mixed Logit (BLP)	36
5.1	Model	37
5.1.1	Latent-class models	37
5.1.2	Random coefficients models	37
5.2	BLP algorithm	38
5.3	DFS algorithm	41

6	Product Availability	42
6.1	Introduction	42
7	Entry Models	43
7.1	Introduction	43
7.2	Static and exogenous models	43
7.2.1	Perfect competition	43
7.2.2	Imperfect competition	44
7.3	Static and endogenous models	45
7.3.1	Sutton's view	45
7.3.2	Bresnahan and Reiss (1991)	45
7.3.3	Berry (1992)	48
7.4	Two-period models	50
7.4.1	Stackelberg's model	50
7.4.2	Empirical work	52

Undergraduate IO Review

- 0.1 Monopoly and Imperfect competition with homogeneous and differentiated products**
- 0.2 Oligopoly with homogeneous products**
- 0.3 Differentiated products**

Chapter 1

Introduction to Empirical Industrial Organization

1.1 Generic Static Analysis

Static analysis refers to the study of firm behavior and market structure in a non-dynamic context. This means that all analysis is conditional on current:

- Goods marketed and their cost functions: investment in output generating units is not considered, nor is R&D, or any other "dynamic elements".
- Consumer preferences: any behavior that might affect changes in preferences is not studied
- Institutional features: changes in the regulation context or market rules are not studied either.

A comparative study will thus involve changing one of these conditions and study the effect on market outcomes by comparison with actual outcomes.

A typical model of static analysis in IO involves three main components:

1. a demand system
2. a cost system
3. an equilibrium assumption

Then, (1) and (2) will be estimated or recovered from actual data, while usually (3) will depend on assumptions made by the economist following knowledge about the institutional setting in which the transactions take place. Generally, economists are very interested in "testing" the assumptions made in (3).

The first class of models that we will cover makes the (strong) simplifying assumption that current decisions of the firms have no effect on future outcomes. This corresponds to a view of markets in which firms are very myopic and do not consider future implications of their choices.

1.1.1 Aside on Dynamic Effects

In comparison with static analysis, dynamic models use state variables that are affected by current and past outcomes and will affect future outcomes. While static analysis recovers immediate impacts, dynamic analysis allows for feedback effects, etc. so that analysis is provided on a longer term.

We will eventually study this type of analysis, but we will cover it following intermediate steps such as two-period models, single-agent dynamic models and then full-fledged dynamic models.

1.1.2 Aside on Theoretical IO and Econometrics

Theoretical advances are not the main drivers of IO research these days. Nevertheless, they still provide essential notions for analysis such as notions of equilibria for different environments and/or institutional settings. This means that a good IO economist should always be able to understand, critique and defend the theoretical models used in the literature or their own work.

A good resource for reviewing modern (yet not complex) IO models is Oz Shy's textbook: "IO: Theory and Applications".

Econometrics provides us the tools of statistic analysis of the features in our models. The econometric side of an IO paper should always include simple descriptive statistics that show a simple correlation to demonstrate a phenomenon,

then a complete analysis of the model used in the research.

1.2 Review of the Cournot model

Suppose the demand system is given by a set of CES demand functions for N different markets, such that:

$$\ln(Q_n) = \beta x_n - \alpha \ln(p_n) + \varepsilon_n$$

where n indexes markets, Q_n is the total quantity sold, x_n is a set of observed market characteristics and ε_n is a set of unobserved market characteristics.

In a Cournot model, firms compete over quantities (Nash-in-quantities). This means that we need to transform the demand curve to an inverse demand curve giving the price for a certain quantity: $p_n(Q_n)$ where $Q_n = \sum_j q_j$ is the sum of all firms' output. We denote firms with the index j and define their cost function as $C_j(\cdot)$ while their marginal cost is mc_j . Firms can have different cost functions and/or different marginal costs.

Thus, the profit function of a single firm (assuming a single market) is:

$$\pi_j(q_j) \equiv p(Q)q_j - C_j(q_j)$$

Assuming full differentiability and strictly positive quantities we get the first-order conditions as:

$$p(Q) + \frac{\partial Q}{\partial q_j} p'(Q) \cdot q_j - mc_j(q_j) = 0$$

In the Cournot setting, firms take the other firms' decision as given, so that they see $\frac{\partial Q}{\partial q_j}$ as equal to 1. This yields:

$$p(Q) + p'(Q) \cdot q_j - mc_j(q_j) = 0$$

and the second-order condition:

$$2p'(Q) + p''(Q) \cdot q_j - mc'_j(q_j)$$

which is strictly negative if and only if:

1.3 Review of the Bertrand model

1.4 Comments on equilibrium notions

1.5 Strategic complementarity of diff. products

Chapter 2

Introduction to Demand Estimation

Recall the multi-product Bertrand model (also found in problem set 0) where a firm f chooses the price vector for its set of products \mathcal{J}_f such that it maximizes its profits. Formally,

$$\max_{\{p_j\}_{j \in \mathcal{J}_f}} \pi_f(\mathbf{p}) = \sum_{j \in \mathcal{J}_f} (p_j - c_j) \cdot q_j(\mathbf{p})$$

where the boldface typed \mathbf{p} represents the vector of all available products (not only those sold by the firm). Taking the FOCs, we get, for each $j \in \mathcal{J}_f$:

$$q_j(\mathbf{p}) + \sum_{k \in \mathcal{J}_f} (p_k - c_k) \cdot \frac{\partial q_k(\mathbf{p})}{\partial p_j} = 0$$

Define the matrix Ω as follows:

$$\Omega_{(j,k)}(\mathbf{p}) = \begin{cases} -\frac{\partial q_k(\mathbf{p})}{\partial p_j} & \text{if } j, k \in \mathcal{J}_f \\ 0 & \text{else.} \end{cases}$$

Then we can rewrite the set of FOCs for all products in \mathcal{J}_f as:

$$q(\mathbf{p}) = \Omega(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{mc})$$

This formula is the most important formula of demand estimation and thus must be understood deeply. To help with understanding it, consider a simple example: there are only $J = 3$ products on the market, of which firm 1 sells only the first two. Its two FOCs are:

$$q_1(\mathbf{p}) = -\frac{\partial q_1(\mathbf{p})}{\partial p_1}(p_1 - c_1) - \frac{\partial q_2(\mathbf{p})}{\partial p_1}(p_2 - c_2) - 0$$

$$q_2(\mathbf{p}) = -\frac{\partial q_1(\mathbf{p})}{\partial p_2}(p_1 - c_1) - \frac{\partial q_2(\mathbf{p})}{\partial p_2}(p_2 - c_2) - 0$$

or in matrix form:

$$\begin{bmatrix} q_1(\mathbf{p}) \\ q_2(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} -\frac{\partial q_1(\mathbf{p})}{\partial p_1} & -\frac{\partial q_2(\mathbf{p})}{\partial p_1} \\ -\frac{\partial q_1(\mathbf{p})}{\partial p_2} & -\frac{\partial q_2(\mathbf{p})}{\partial p_2} \end{bmatrix} \cdot \left(\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} - \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right)$$

but we miss the third product, so in order to include it we add a line. The final matrix form is given by:

$$\begin{bmatrix} q_1(\mathbf{p}) \\ q_2(\mathbf{p}) \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{\partial q_1(\mathbf{p})}{\partial p_1} & -\frac{\partial q_2(\mathbf{p})}{\partial p_1} & 0 \\ -\frac{\partial q_1(\mathbf{p})}{\partial p_2} & -\frac{\partial q_2(\mathbf{p})}{\partial p_2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} - \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \right)$$

$$\Leftrightarrow q(\mathbf{p}) = \Omega(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{mc})$$

As stated before, this equation is the most important equation of demand systems. Once we estimate (or assume partly) predicted sales $q(p)$ and elasticities Ω , we can recover markups and marginal costs. This will be particularly useful for recovering data that we do not have and completing further analysis of the market such as: comparative statics, welfare impacts, research on advertising, research on elasticities, evaluation of new products, etc.

Nevertheless, demand estimation is not that easy. In fact, a demand system incorporates two major challenges that make it particularly hard to estimate: the endogeneity of prices and the flexibility need for multiproduct demand.

Endogeneity of prices arises because prices are not randomly assigned to quantity levels, but they are set as a result of a strategic decision which typically include quantity considerations. This means that there exists simultaneity in the process of coming up with equilibrium prices and quantities. Different market

outcomes could be linked to either demand shifts, supply shifts or both, without the econometrician observing it. This issue is mainly addressed by using instruments.

The flexibility issue is linked to the dimensionality of the matrix Ω . As the number of products J increases, the number of own- and cross-price elasticities increase quadratically (there are J^2 elasticities to estimate). Moreover it might also be the case that the elasticities matrix varies with p in a way that is not captured by the model (think about $\partial \mathbf{p} / \partial p_j$). Finally, one could consider what can be said of second-order derivatives or third-order derivatives, increasing the dimensionality even more.

There are many types of demand models, but most of them can be separated in these categories:

- Representative vs. Heterogenous consumers?
- Discrete vs. continuous choice?
- Single vs. multiple products?
 - Within the multiple products case: products vs. characteristic space?

The data used in these types of models are often aggregate, market-level data on the pairs (p_j, q_j) . Sometimes it contains more details to it such as the characteristics of the goods, the characteristics of the individuals buying the goods (either aggregate or micro-data), the environment around, etc.

2.1 An early example: Bresnahan (1987)

Bresnahan (1987) is one of the first empirical assessment of competition using demand estimation, spurring the birth of what is now called NEIO (New Empirical IO). The paper studies the presence of collusion in the automobile industry by looking at a dramatic price decrease event that happened in 1955.

The intuition of this paper is quite simple: assuming marginal costs stayed relatively constant during that period, Bresnahan estimates the variation in demand elasticities and compares it to a change in competition model. Using data on 85 models over 3 years, he finds that a change from a collusive to a more competitive equilibrium is consistent with the estimated change in elasticities.

2.2 Single product demand

2.2.1 Representative agent

A demand system is generally thought to be a (more or less) general approximation of a true underlying demand curve, for example:

$$\ln(q) = \alpha p + X\beta + \varepsilon$$

where the ε captures any variation that is unobserved by the econometrician. Using that equation, the objective of the econometrician is to recover the underlying parameters (α, β) in order to understand the effects of prices and characteristics on demand. Note that this specification assumes a representative consumer choosing a (total) quantity q for a given price p . On the other side, the firms that provide the good might also choose their prices p and characteristics x as a function of the expected demand, which is correlated with q . If this is the case, we say that we have endogeneity (from simultaneous equations, as we've seen in econometrics). Since the coverage of IV estimation in that case is more of an econometrics topic, we will assume knowledge of this in the rest of the chapter.

2.2.2 Heterogeneous agents

While the previous model is very simple, a straightforward extension would be to allow for heterogeneous agents on the market. To do that, we need to extend the previous model in two ways: (1) use a micro-founded model for individual demand that aggregates nicely and (2) estimate aggregated demand as:

$$\ln(q) = \int \gamma_i g(\gamma) d\gamma + \int \alpha_i p f(\alpha) d\alpha + \int \beta_i x h(\beta) d\beta + \epsilon$$

where α, β and γ follow known distributions with unknown parameters to be estimated.

2.3 Multi-product demand

2.3.1 Product Space approach

This course explores product space approaches only briefly since most of modern IO revolves around the characteristic space approach.

In product space, the basic breakdown between models is whether they include a representative agent or heterogeneous agents. The most interesting of them both is the heterogeneous agents ones since it uses more of the available information (better estimates) but also provides a framework for distributional policy analysis (estimating the distribution of responses). These models have been around for a long time but they still face big issues such as treatment of zeroes (when some consumers have no access to a particular good) or aggregation issues (computational). This is why much of the work that has been done in product space revolves around representative agent models, starting with the AIDS model by Deaton and Muellbauer (1980).

Almost Ideal Demand System (AIDS)

The motivating goal behind the AIDS model is to get a very general, aggregate demand system starting from consumer theory. This demand system is one of the first to exactly aggregate over consumers. This will provide a system that will provide the Ω matrix that we are looking for, while satisfying the basic axioms of consumer theory.

Begin by defining a general expenditure function for a given group of (comparable) goods:

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

where u is given between 0 (subsistence) and 1 (bliss). This means that $a(\mathbf{p})$ is the cost of subsistence and $b(\mathbf{p})$ is the cost of bliss. The expenditure function therefore represents a linearly homogenous function of both costs.

Now, in order to go deeper, we need to specify exactly what are $a(p)$ and $b(p)$. Intuitively, $\ln b(p)$ should include $\ln a(p)$ since it is the cost of going beyond

subsistence. In particular we will have:

$$\ln a(p) = a_0 + \sum_k a_k \ln(p_k) + (1/2) \cdot \sum_k \sum_j \gamma_{jk}^* \ln(p_k) \ln(p_j)$$

$$\ln b(p) = \ln a(p) + \beta_0 \prod_k p_k^{\beta_k}$$

This yields the complete expenditure function as:

$$\ln e(u, p) = a_0 + \sum_k a_k \ln(p_k) + (1/2) \cdot \sum_k \sum_j \gamma_{jk}^* \ln(p_k) \ln(p_j) + u \cdot \beta_0 \prod_k p_k^{\beta_k}$$

We estimate $(\alpha_i, \beta_i, \gamma_{jk}^*)$ from the data, usually under the following assumptions:

- $\sum_i a_i = 1$:
- $\sum_j \beta_j = 0$:
- $\sum_k \gamma_{jk}^* = 0$:
- $\gamma_{jk}^* = \gamma_{kj}^*$:

Applying Shepard's Lemma and log differentiation, we can obtain the expenditure share of any good k within the group, denoted w_k :

$$w_k = a_k + \sum_j \gamma_{kj} \ln p_j + \beta_k \ln(x/P)$$

where γ_{jk} is the average of γ_{jk}^* and γ_{kj}^* ; x is the total expenditure within the group and P the price index of the group.

There are two ways of computing the price index of the group:

- Directly from the specification:
- As a linear approximation:

2.3.2 Issues with AIDS

All in all, AIDS works very well when products are actually grouped in certain categories. This is the reason why it is widely used in the trade and macro-consumption literatures. It can be used in IO settings where groups could be single products but it requires instrumenting for every good!

Think of the case where there are J goods, then you need to estimate J^2 elasticities or at least $J \times (J + 1)/2$ if you are assuming that cross-price elasticities are symmetric. Usually, IO economists try to escape the problem by actually grouping the goods, however all the results will be dependent on your initial choice of grouping. Researchers have to think very deeply in the mechanism they use for grouping. Grouping subject to particular characteristics lead you closer to characteristic space models, but even then, how do you group when the characteristics are continuous etc.

Finally, let's look at one of the main challenges with the product space approach. Suppose a new good is introduced in a market, how can you use the previous demand estimated which did not account for the new product? This creates two issues:

1. Equilibrium effects: the introduction of a new good typically has significant effects on the pricing of the previously existing goods.
2. Extrapolation: using the previously estimated demand implies projecting demand where it is not defined.

2.3.3 Examples: Hausman, Leonard and Zona (1994)

The goal of the paper is to estimate demand for beer in the US in order to perform a merger analysis and test assumptions about the firms' conduct.

The market is divided hierarchically in three levels:

1. The top-level is beer against other goods. This is estimated using a log-log expenditure function:

$$\ln e_t = \beta_0 + \beta_1 \ln y_t + \beta_2 \ln P_{bt} + Z_t \delta + \varepsilon_3$$

where e_t is total expenditures, y_t is income and P_{bt} is the beer price index.

2. Within the beer group, the middle-level are the segments: premium, popular and light. Again, a log-log functional form is assumed so that for a given segment m , the demand for the segment is:

$$\ln q_m = \beta_m \ln e_t + \sum_{m'} \sigma_{m'} \ln \pi_{m'} + \alpha_{m'} + \varepsilon_2$$

3. Within each segment, the bottom-level contains five brands. This level is assumed to follow an AIDS:

$$w_k = a_k + \sum_j a_{jk} \ln p_j + \beta_k \ln(x/P) + \varepsilon_1$$

In the bottom-level, you have to instrument the price since it is correlated with unobserved product quality (taste, etc.) and unobserved demand shocks (special events like the World Cup, etc.). Then you must find brand-level instruments: this is where the infamous Hausman instruments come in. Hausman instruments use prices in one city to instrument for prices in other cities. These instruments are typically strong, however, their relevance is very questionable. In particular, think about nation-wide advertising which was the main criticism of these instruments. They would clearly affect both cities' prices in the same way creating correlation. Most of the time you can observe these patterns but be wary of unobserved shocks that could affect the complete market. Optimally, the best instrument would be observed input costs for the brand, tax rates, etc.

Hausman results:

2.3.4 When should you choose product space approach?

2.4 Multi-product demand: Characteristic Space

The characteristic space approach follows a different philosophy than the product space:

- Products are considered as a vector of characteristics
- Consumers' preferences (and thus their utility functions) are defined over these characteristics.
- Consumers are assumed to choose the characteristic vector (the product) that maximize their utility.
 - One consumer makes one choice to buy a single product. Allowing for buying more than one product is computationally costly and is an open area for research.
- Demand is aggregated by simply summing over consumers' choices.

2.4.1 History of the characteristics approach

2.4.2 Formal base model

Assume the following context:

- A consumer i is offered J alternatives. Consumer i is identified by his characteristics, v_i , similarly to the product being defined by its characteristics x_j and its price p_j .
- He must choose one option only $j \in J$, which he'll do with probability P_{ij}
- Utility of an individual i for a good j is given by:

$$U_{i,j} \equiv U(x_j, p_j, v_i, \theta) \text{ for all } j = 0, 1, \dots, J$$

where good $j = 0$ is referred to as the outside good.

The variable (or most probably the vector) x_j contains non-price characteristics about good j , for example the size of the engine, the AC system, the dashboard software, etc.; while v_i is the vector of consumer characteristics such as income, age, etc. Finally, θ is the vector of parameters of the utility which is to be estimated. Note that in order to estimate the model correctly, you need to be sure that all consumers (or groups of consumers) in the sample have the same access to all J goods, or in other words, they face the same choice set.

Consumer i will choose good j if and only if $U_{i,j} > U_{i,k}$ for any other good k . This means that the probability of consumer i buying good j is given by:

$$\Pr [U_{ij} > U_{ik} \text{ for all } k \neq j]$$

Now assuming that utility is given by an observable component V_{ij} and an unobservable component ε_{ij} , we can rewrite the probability of buying the good as:

$$\begin{aligned} P_{ij} &= \Pr [U_{ij} > U_{ik} \text{ for all } k \neq j] \\ &= \Pr [V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik} \text{ for all } k \neq j] \\ &= \Pr [\varepsilon_{ij} - \varepsilon_{ik} > V_{ik} - V_{ij} \text{ for all } k \neq j] \end{aligned}$$

Finally, define $f(\varepsilon_i)$ as the J -dimensional probability density function for consumer i . Then the probability P_{ij} can be written as:

$$P_{ij} = \int \mathbb{I}\{\varepsilon_{ij} - \varepsilon_{ik} > V_{ik} - V_{ij}\} f(\varepsilon_i) d\varepsilon_i$$

which is (as ε_i) a J dimensional integral over $f(\varepsilon_i)$.

We'll see later on that some choices could make our life easier in estimating this integral (hard to do it analytically), mainly by parameterizing the error term:

- **Multivariate normal:** $\varepsilon_i \sim N(0, \Omega)$ → multinomial probit model.
- **Type 1 Extreme-value:** $f(\varepsilon_i) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}}$ → multinomial logit model.
- other variants...

We could also turn to thinking about absolute demand for good j instead of its probability, which is given by the size of the set $S_j(\theta)$ defined as:

$$S_j(\theta) \equiv \{i : U_{i,j} > U_{i,k} \text{ for all } k\}$$

Now suppose consumers are distributed following a pdf $f(v|\theta)$, we could recover the market share of good j as:

$$s_j(\mathbf{x}, \mathbf{p}|\theta) = \int_{i \in S_j(\theta)} f(v) dv$$

Given a total market size of M , total demand for good j is: $M \cdot s_j(\mathbf{x}, \mathbf{p}|\theta)$

As we've seen the complete model of characteristic space relies on utility functions and error terms distributions. This is why, unsurprisingly, the functional forms of utility and errors are exactly what will differentiate the models within the characteristic space approach.

2.4.3 Types of models

There are 5 main models of characteristic space.

Pure horizontal model

Analog to the Hotelling model (in its simplest form), there are n ice cream sellers along a beach where consumers are distributed. A consumer i has utility:

$$U_{ij} = \bar{u} - p_j + \theta(\delta_j - v_i)^2$$

where the term $(\delta_j - v_i)^2$ captures some kind of quadratic transportation cost in the distance from i 's preferences (v_i) to the characteristics of product j (δ_j). In the simple Hotelling model, δ_j and v_i are location parameters (where ice cream seller j and consumer i are on the beach).

Pure vertical model

On the pure vertical, you don't need to model a distance from consumer preferences since everyone agrees on what product is better. In this case, utility is given by:

$$U_{ij} = \bar{u} - v_i p_j + \delta_j$$

The interaction between consumer characteristics and prices show that although every consumer has the same utility for the characteristics of good j (i.e. $\bar{u} + \delta_j$), consumers differ in their willingness to pay (v_i).

Logit model

In the logit model, consumers have the same taste for the goods' characteristics but are subject to an idiosyncratic shock depending on both the product and the consumer. Utility is given by:

$$U_{ij} = \delta_j - p_j + \varepsilon_{ij}$$

where ε_{ij} follows an iid extreme-value type I distribution. Analogous to the OLS, the error term allows for identification of the other parameters as long as it is not correlated with the δ_j and the price.

Pure characteristic model

This model nests both concepts of verticality and horizontality of differentiation.

Utility is given by:

$$U_{ij} = f(v_i, p_j) + \sum_k \sum_r g(x_{jk}, v_{ir}, \theta_{kr})$$

so that it could handle vertical model if $g(\cdot)$ is 0 for cross-products variables. It could also handle the horizontal model. It is a very general form and rarely used?

BLP model

The BLP model (for the authors' names Berry, Levinsohn and Pakes), is a parameterized version of the pure characteristics model. It is probably the most commonly used demand model in the empirical literature as of now, and is also very popular in non-research applications of IO. Utility is given by:

$$U_{ij} = f(v_i, p_j) + \sum_k \sum_r x_{jk} v_{ir} \theta_{kr} + \varepsilon_{ij}$$

where the ε_{ij} extreme-value Type I error term has been added.

2.4.4 Data and issues with estimation

Chapter 3

Estimating demand for vertically differentiated products

Chapter 4

Logit, Nested Logit and Multinomial Probit

4.1 "Plain-vanilla" Logit

4.1.1 Overview

Recall that in the pure logit model, each product has a mean quality level δ_j that incorporates the characteristics of the product as well as its price, and an idiosyncratic error term ε_{ij} that is drawn iid, from a Type-I Extreme Value (or Gumbel) distribution. The Gumbel distribution has a pdf of $f(x) = e^{-x}e^{-e^{-x}}$ and a cdf of $F(x) = e^{-e^{-x}}$. We assume the variance of that distribution to be equal to $\pi^2/6$ which will normalize the scale of utility as we'll see later. The mean of the distribution is non-zero but we'll see shortly why it is not an issue. Formally, the utility derived from consumption of a good j by consumer i is written as:

$$U_{ij} = \delta_j + \varepsilon_{ij}$$

where $\delta_j = X_j\beta - \alpha p_j + \xi_j$. We say that a consumer chooses a good j over k if and only if his utility from good j is strictly higher than from good k .

Therefore, the probability that consumer i buys good j is the probability that he

chooses j over any k :

$$\begin{aligned} P_{ij} &= \Pr [U_{ij} > U_{ik} \text{ for all } j \neq k] = \Pr [\delta_j + \varepsilon_{ij} > \delta_k + \varepsilon_{ik} \text{ for all } j \neq k] \\ &= \Pr [\varepsilon_{ij} - \varepsilon_{ik} > \delta_k - \delta_j \text{ for all } j \neq k] \end{aligned}$$

It turns out that if ε_{ij} and ε_{ik} follow a type-I extreme value distribution independently, then the difference between both follows a logistic distribution with cdf $F(x) = \frac{e^x}{1+e^x}$. This form of the probability also tells us that constant terms in the utility across products do not matter since we consider only $\delta_j - \delta_k$ which exactly cancels any constant that we could observe. Using this fact, we can normalize utility by subtracting a same constant to all values, in particular, we can set the utility of an outside good to 0, so that $\delta_0 = 0$.

Since all error terms are independent of each other, we can write the following:

$$\begin{aligned} P_{ij}|\varepsilon_{ij} &= \Pr [\varepsilon_{ik} < \varepsilon_{ij} + \delta_j - \delta_k \text{ for all } j \neq k] \\ &= \prod_{k \neq j} F(\varepsilon_{ij} + \delta_j - \delta_k) \end{aligned}$$

and since ε_{ij} is not known, we compute the unconditional probability by integrating over all possible values of ε_{ij} from the Gumbel distribution:

$$P_{ij} = \int \prod_{k \neq j} F(\varepsilon_{ij} + \delta_j - \delta_k) f(\varepsilon_{ij}) d\varepsilon_{ij} = \frac{e^{\delta_j}}{\sum_{k=0}^J e^{\delta_k}}$$

You can see that the probability of consumer i buying good j does not depend on any individual-only parameter or variable, so that we can express this probability of buying a good as the market share of that good s_j . Moreover, since we assume that utility for the outside good was 0, we have that $e^{\delta_0} = 1$. This yields the following market shares:

$$\begin{aligned} s_j(\theta) &= \frac{\exp(\delta_j)}{1 + \sum_k \exp(\delta_k)} \\ s_0(\theta) &= \frac{1}{1 + \sum_k \exp(\delta_k)} \end{aligned}$$

This model is very practical in the sense it is very easy to set up and displays several desirable properties:

- $s_j = P_{ij} \in (0, 1)$ so that any market share can be rationalized by the model.
- As δ_j increases, reflecting higher attributed quality, the market share will increase to 1. If δ_j decreases, then the market share goes to 0. However, note that these results only hold at the limit, meaning a good will never have a market share equal to 0 or 1. If that is the case in the data, the researcher should take the product out of the dataset. This is why estimating a logit model must be done on a time-period long enough so that all goods considered have been sold at least once (think about vending machines).
- $\sum_j s_j = 1$, meaning that one alternative will be chosen, always.
- $s_j = P_{ij}$ is everywhere continuously differentiable in the characteristics δ_j and in price p_j .

The last point is useful to compute price derivatives (and elasticities):

$$\begin{aligned}
\frac{\partial s_j(\theta)}{\partial p_j} &= \frac{-\alpha e_j^\delta \cdot (1 + \sum_k e_k^\delta) + \alpha e_j^\delta e_j^\delta}{(1 + \sum_k e_k^\delta)^2} = \frac{-\alpha e_j^\delta \cdot [1 + \sum_k e_k^\delta - e_j^\delta]}{(1 + \sum_k e_k^\delta)^2} \\
&= -\alpha \cdot \frac{e_j^\delta}{1 + \sum_k e_k^\delta} \cdot \frac{1 + \sum_k e_k^\delta - e_j^\delta}{1 + \sum_k e_k^\delta} \\
&= -\alpha s_j(1 - s_j)
\end{aligned}$$

And using the same method, we find $\partial s_j(\theta)/\partial p_k = \alpha s_j s_k$

4.1.2 Logit and taste variation

The logit model is very efficient and easy to set up if you need to measure systematic taste variation, that is, variation in tastes associated with observable characteristics X_j . Variations associated with idiosyncratic randomness, ε_{ij} are assumed away as type-I extreme value random variables.

In reality, the value that agents put on a particular attribute varies across these individuals. Sometimes these tastes vary in an identifiable way, for example, low-income agents might more concerned about the price than others; sometimes you might observe people with the exact same observable characteristics choosing different options. Logit models allow for identification of these taste variations only within limits: if this variation is linked with observable characteristics, then

logit models can be applied; if this variation is purely random, then we will require other models.

As an example, consider the choice of households between cars, where two characteristics enter the decision: price p_j and shoulder room x_j . A household i places value U_{ij} on buying good j such that:

$$U_{ij} = \beta_i x_j - \alpha_i p_j + \varepsilon_{ij}$$

where you can see that the parameters α, β vary across households i . Say the shoulder room taste β_i depends only on the number of members in the household m_i such that $\beta_i = \rho M_i$; this means that m_i is positively correlated with β_i . Similarly let's say the importance of price is negatively correlated to income I so that $\alpha_i = \theta/I_i$. Substituting back into the utility function we get:

$$U_{ij} = \rho M_i x_j - \theta p_j / I_i + \varepsilon_{ij}$$

which can be estimated with a standard logit model where variables are interaction between characteristics of the product and characteristics of the household.

In contrast, suppose taste variation was subject to an error term such that $\beta_i = \rho M_i + \mu_i$ where μ_i is non-observable (a random variable). Then the utility would be written as:

$$U_{ij} = \rho M_i x_j - \theta p_j / I_i + \varepsilon_{ij} + \mu_i x_j$$

and the logit would confound the error term with $\tilde{\varepsilon}_{ij} = \varepsilon_{ij} + \mu_i x_j$ which is clearly correlated across product specifications and households.

Bottomline is logit models can handle systematic taste variation but not random taste variation. For the latter, we will need more complex models, such as the BLP model studied in the next chapter.

4.1.3 Too-many-parameters problem

The estimation of the quality characteristic δ_j is done over the set of non-price characteristics X_j . In particular, recall that we've seen that:

$$\delta_j = X_j \beta - \alpha p_j + \xi_j$$

In order to identify correctly this quality term, we need that the number of different observed characteristics (the dimension of X_j , say k) is smaller than the number of products j . This also means that we will need to include the term ξ_j in our regression in order to explain any pattern of market shares that would be unobserved by the econometrician.

However, if consumers know ξ_j , firms must also know it and price their products accordingly. This introduces endogeneity in the error term with respect to prices. Thus, the econometrician needs instruments to correct for that endogeneity. Typically, these instruments include cost shifters that would not affect intrinsic demand for the good but would definitely have an effect on prices.

Note that any non-price characteristic included in X_j that would also be correlated with the unobservable characteristics ξ_j needs to be instrumented as well.

4.1.4 Derivatives and Elasticities

The logit model presented above displays features that are not desired in the context of elasticities. These problems imply that one would run into issues while trying to expand logit results in terms of welfare analysis, antitrust analysis, etc.

First, recall the own- and cross-price derivatives and elasticities:

$$\frac{\partial s_j}{\partial p_j} = -\alpha s_j(1 - s_j); \quad \frac{\partial s_j}{\partial p_k} = \alpha s_j s_k; \quad \epsilon_{jj} = -\alpha p_j(1 - s_j); \quad \epsilon_{jk} = \alpha p_k s_k$$

These imply that:

- Two products with the same market shares will have the same markups. Indeed, from the first-order conditions, we have that:

$$p_j - mc_j = \frac{1}{|\epsilon_{jj}|} \Leftrightarrow \frac{p_j - mc_j}{p_j} = \frac{1}{\alpha(1 - s_j)}$$

which would be the same for $s_j = s_k$. This is obviously not intuitive but also not observed in reality.

- Own-price elasticities are higher for higher priced goods. This fact comes directly from the formula of the own-price elasticity that shows a positive

effect (in absolute value) from the prices. This is counter-intuitive since it would imply that people buying higher-priced items would be more price-sensitive than people buying lower-priced goods.

- Substitution between goods only depends on relative shares and not proximity of product characteristics. In fact, cross-price elasticities are given by $\frac{\partial s_j}{\partial p_k} \frac{p_k}{s_j} = \alpha p_k s_k$, which is not a function of characteristics of neither products.

4.1.5 Independence of Irrelevant Alternatives (IIA)

We have seen in the overview of the logit model that idiosyncratic errors are independently distributed across products, following a type-I extreme value distribution. The choice of the distribution turns out not to be very important, but the independence property has some implications that the researcher should know about. In fact, the independence assumption means that the unobserved utility derived from one good (ε_{ij}) is unrelated unconditionally to the unobserved utility from another good. This assumption seems to be rather restrictive in the context of goods but helped us a lot in coming up with the solution of the logit model.

The independence of the error terms turns out to have problematic implications in the realism of the choice mechanism. In fact, suppose a consumer has a choice of going to work by car (c) or taking a blue bus (bb). Say that utility derived from both models are the same, such that $P_c = P_{bb} = 1/2$, meaning the ratio of probabilities is one. Now suppose that a red bus (rb) is introduced in the market such that it is identical in every aspect except the color to the blue bus. The probability of taking the red bus should therefore be the same as taking the blue bus: $P_{rb}/P_{bb} = 1$. However, the ratio of probabilities between the car and the blue bus has not changed because of independence of irrelevant alternatives, meaning that $P_{bb}/P_c = 1$, thus $P_c = P_{bb} = P_{rb} = 1/3$ is the prediction of the logit model. In real life though, we would expect the probability to take the car to remain exactly the same since the actual problem is to either take the bus (regardless of the color) or the car, yielding $P_c = 1/2, P_{bb} = P_{rb} = 1/4$.

This IIA problem is nonetheless also a feature of the logit model when it corresponds to reality. In fact, this property allow the researcher to consider only

subsets of the complete set of alternatives and still get consistent estimates, as long as for each observation, the actual choice is kept in the set. Another practical use of that property is that if the researcher is only interested in a few choices, then they do not need to include the other choices in the dataset, leaving the data research part out of the picture.

One could test for these properties by, ...

4.1.6 Consumer Surplus

For policy analysis, the researcher is often interested in measuring the change in consumer surplus that is associated with a particular event (introduction of a product, merger, etc.). Under the logit assumption, this value of the consumer surplus also takes a simple closed form that is easily found from the model. We know that each consumer will choose the product that yields the highest utility. In the aggregate case, all consumers have the same tastes so that in expectation (the econometrician does not observe ε_{ij}), consumer surplus is defined as the value of utility derived from the best good. Formally,

$$E[CS] = E \left[\max_j \{U_{ij}\} \right] = E \left[\max_j \{\delta_j + \varepsilon_{ij}\} \right]$$

which yields:

$$E[CS] = \ln \left(\sum_{j=1}^J e^{\delta_j} \right) + C$$

where C is an unknown that represents the fact that absolute utility cannot be measured. This value is called the logit inclusive value, or the log-sum term. As you can see, comparing two policies is easy in this setting, let one policy be denoted by the superscript 0 and the other by the superscript 1. Then,

$$\Delta E[CS] = \ln \left(\sum_{j^1} e^{\delta_j^1} \right) - \ln \left(\sum_{j^0} e^{\delta_j^0} \right)$$

If we had transaction-level data with individual characteristics, we could perform this analysis of consumer surplus at each unit of observation and aggregate to find our results.

4.2 Nested Logit

A natural extension of the simple logit model, allowing for richer substitution patterns and a somewhat less restrictive IIA property is the nested logit model.

A nested logit model partitions the choice set in different subsets called nests such that the actual choice of one good follows from choices among nests. Consumers choose a category of products (modelled as a simple logit), then within a category, they choose a product (again modelled as a simple logit). This sequence of logit model creates a different type of model called the nested logit, where the nests are the products categories. These nests are chosen by the researcher (which is not ideal) and they provide a strong structure to the model which could affect results significantly. Following the notation of [?] and [?], we model utility as:

$$U_{ij} = \delta_j + \zeta_{ig} + (1 - \sigma)\varepsilon_{ij}$$

where the new terms are: ζ_{ig} an idiosyncratic "nest" taste shock that applies to all goods j in the nest g ; and σ a parameter of correlation in tastes within nests (if σ is high, tastes within group are very correlated and the nest structure matters, if σ is low, then the correlation in tastes within the nest is small and the nest structure is irrelevant). The ζ_{ig} error term in this utility follows the unique distribution distribution function such that $\zeta_{ig} + (1 - \sigma)\varepsilon_{ij}$ follows an extreme-value distribution (not type-I however, we call it Generalized Extreme Value or GEV).

We know that within the same nest, the utility level $u_{ij} = \delta_j + \zeta_{ig}(1 - \sigma)\varepsilon_{ij}$, can be reduced down to ignore the effect of ζ , since it is the same across products. We end up in the same setting as the simple logit model. This implies that the conditional share of product j , or the share within the nest, is given by the same formula as the simple logit:

$$s_{j|g} = \frac{\exp(\delta_j/(1 - \sigma))}{\sum_{k \in \mathcal{J}_g} \exp(\delta_k/(1 - \sigma))}$$

where the denominator for this expression can be written as D_g , the total demand for products in the group g .

Meanwhile, across groups, we have that both error terms still give a type-I EV, so

again, we can think about the demand for a group as we did in the logit case:

$$s_g = \frac{D_g^{(1-\sigma)}}{\sum_h D_h^{(1-\sigma)}}$$

Finally, the share of a product j is given by the product of both the share of the group containing j and the conditional share of j within the group:

$$s_j = s_{j|g} \cdot s_g = \frac{\exp(\delta_j/(1-\sigma))}{D_g^\sigma \cdot \left(\sum_h D_h^{(1-\sigma)}\right)}$$

As we can see, demand for product j depends on its own quality level relative to its group, the quality of the group relative to the other groups and σ , the correlation in tastes within nests.

The outside good in this model is considered as one of its own group, so that $s_{0|0} = 1$ and with the normalization of $\delta_0 = 0$ and $D_0 = 1$, we get:

$$s_0 = 1 \cdot s_0 = \frac{1}{\sum_h D_h^{(1-\sigma)}}$$

This analytical derivation is the same when we extend the model to have nests within nests and more.

4.2.1 IIA and substitution patterns

The nest structure of this model satisfies two properties:

1. For any two alternatives that are in the same nest, the ratio of probabilities is independent of the attributes and/or existence of other products in the nest or in other nests.

$$s_j/s_k = s_{j|g}/s_{k|g} = \frac{\exp(\delta_j/(1-\sigma))}{\exp(\delta_k/(1-\sigma))}$$

Equivalently, we say that IIA holds within the nest.

2. For any two alternatives that are in different nests, the ratio of probabilities depends on the attributes and/or existence of other products in the two nests.

$$s_j/s_k = s_{j|g}/s_{k|h} \cdot s_g/s_h = \frac{\exp(\delta_j/(1-\sigma))}{\exp(\delta_k/(1-\sigma))} \cdot \frac{D_g^{1-\sigma}}{D_h^{1-\sigma}}$$

Equivalently, we say that IIA does not hold across nests.

4.2.2 Research considerations

As we have seen, the nested logit is a fairly simple extension on the simple logit case, which relies on relaxation of the IIA property across groups (or nests). The model is often told in a narrative of sequential choice: consumers first choose a group g (s_g), then a product (or group) j within group g ($s_{j|g}$).

The results of the model depend very strongly on the ex ante nesting structure chosen by the researcher. Therefore, it is very important (and hard) to understand what the appropriate structure should be. The sequential narrative is supposed to help the researcher come up with an accurate structure but it might be unhelpful at times.

Identification comes in different flavors:

- Parameters associated with characteristics and prices are identified within group by variation in these exact characteristics.
- The correlation in tastes parameter (σ) is identified by variation in

4.2.3 Extensions of the GEV

4.3 Multinomial Probit

The multinomial probit is a more flexible alternative to the simple logit model. In fact, by allowing errors ε_{ij} to be correlated (not iid) and to follow a multivariate normal distribution $N(\mu, \Sigma)$ that is completely free, it achieves more interesting

outcomes than the simple logit. For example, with an unrestricted variance matrix Σ , you could end up with substitution patterns that are more...

4.4 GMM Estimation Strategy for Product-level data

The general estimation strategy that applies to logit models with product-level data is detailed in the following steps:

1. Assume that data is drawn from markets with large n .
2. Assume that observed market shares are measured without errors.
3. For each θ (vector of parameters), there exists a unique ξ such that the model shares and observed shares are equal (J equations, J unknowns).
4. We invert the model to get ξ as a function of the parameters. How this step is performed depends on the functional form of the model.
5. Using ξ , we can create the moments of the model, estimating them by GMM.

4.4.1 Inversion in the logit case

We need to express the quality unobservable term ξ_j in terms of all observable characteristics. First, recall that:

$$\delta_j = X_j\beta - \alpha p_j + \xi_j$$

but in this equation, δ_j is unknown (there is no such thing as a perceived mean quality level). Nevertheless, we can use the market shares formulas that link the δ_j to the observed market shares s_j (observed without errors). Thus,

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_k \exp(\delta_k)}$$

but again, we run into a problem since it is expressed as a function of other δ_k : we could solve a system of equations, or simplify the model using the normalized good. Indeed,

$$\frac{s_j}{s_0} = \exp(\delta_j) \Leftrightarrow \ln(s_j) - \ln(s_0) = \delta_j \Leftrightarrow \ln(s_j) - \ln(s_0) = X_j\beta - \alpha p_j + \xi_j$$

which in turn yields the inversion equation:

$$\ln(s_j) - \ln(s_0) + \alpha p_j - X_j \beta = \xi_j$$

4.4.2 Inversion in the nested-logit case

In the same way as in the simple logit model, our main objective is to get the link between δ_j and observables so that we can identify ξ_j . In this case:

$$s_j = \frac{\exp(\delta_j/(1-\sigma))}{D_g^\sigma \cdot \left(\sum_h D_h^{(1-\sigma)}\right)} \text{ and } s_0 = \frac{1}{\sum_h D_h^{(1-\sigma)}}$$

so that

$$\ln(s_j) - \ln(s_0) = \frac{1}{1-\sigma} \cdot \delta_j - \sigma \ln(D_g)$$

We turn into a new problem caused by the two-level structure, which is D_g is not parameterized. Again, we solve this issue by using the normalized good:

$$s_g/s_0 = D_g^{(1-\sigma)} \Leftrightarrow \ln(s_g) - \ln(s_0) = (1-\sigma) \cdot \ln(D_g)$$

which we can plug back in the previous equation to get:

$$\begin{aligned} \ln(s_j) - \ln(s_0) &= \frac{1}{1-\sigma} \cdot \delta_j - \frac{\sigma}{1-\sigma} [\ln(s_g) - \ln(s_0)] \\ \Leftrightarrow (1-\sigma) \cdot [\ln(s_j) - \ln(s_0)] &= \delta_j - \sigma [\ln(s_g) - \ln(s_0)] \\ \Leftrightarrow (1-\sigma) \cdot \ln(s_j) + \sigma \ln(s_g) - \ln(s_0) &= \delta_j \end{aligned}$$

Finally, using the fact that $s_g = s_j/s_{j|g} \Leftrightarrow \ln(s_g) = \ln(s_j) - \ln(s_{j|g})$, we can get:

$$\Leftrightarrow \ln(s_j) - \sigma \ln(s_{j|g}) - \ln(s_0) - X_j \beta + \alpha p_j = \xi_j$$

4.4.3 Endogeneity issues

Regardless of the model we use, we will have to deal with endogeneity issues regarding many dimensions, such as prices, observable characteristics or market shares.

In fact, in both models price was correlated to the unobservable term as firms make their decision based on demand which includes this term; moreover, in the nested logit case, we also have to deal with endogeneity in the within-share term. In general, anything that is correlated with the unobservable quality term ξ will have to be instrumented.

In the particular case of the nested logit model, you should realize that conditional market shares and actual market shares depend on the same "function", which means we also need to instrument for one of these shares. ...

4.4.4 Measurement errors

Measurement errors in observed prices, characteristics or quantities may also create difficulties for the estimation procedure outlined above. Prices enter linearly in the estimation and are already endogenous so that these errors do not create too important biases. However, when these errors are present in market shares or quantities data then this issue becomes more important. Indeed, market share data is used to invert the model and get the unobservable term as a function of observed data. This non-linear inversion will in fact be very sensitive to measurement errors in the market shares.

4.4.5 Adding supply-side restrictions

4.5 Individual-level data

Chapter 5

Mixed Logit (BLP)

For now we have seen three types of discrete-choice models and their applications to demand estimation for differentiated products: the simple logit model, the nested logit model and the multinomial probit model. These three models, although quite useful and fairly simple to estimate were limited in three dimensions: they do not allow for random taste variation (4.1.2), they have restricted substitution patterns (4.1.5), they do not allow for correlation over time. Mixed logit models (which contains BLP) are highly flexible models that can deal with the previously mentioned issues.

Mixed logit models are a class of models that encompasses all types of models where the market shares are computed as integrals over simple logit functional form. We'll see in detail later what that means but intuitively, you should think of the model as everyone having her/his own logit model of demand, and aggregate demand would be computed by integrating over consumer attributes. In particular, IO economists are most interested in the BLP (for [?]) model and extensions of it like [?].

We'll first cover the basics of mixed logit and random coefficients, before talking in depth about estimation techniques as found in BLP and DFS.

5.1 Model

Generally, as stated in [?], we write utility derived from consumption of a good j by consumer i as the function $u_{ij}(X_j, p_j, \xi_j, \nu_i, \theta)$, which is a function of product j observable characteristics (X_j), price (p_j), unobservable characteristics (ξ_j) and consumer characteristics (observable z_i and unobservables ν_i), all entering the utility function through a vector of parameters θ .

As a simpler case, define utility as a linear function of those parameters such that:

$$U_{ij} = X_j\beta_i + \xi_j + \varepsilon_{ij}$$

Notice the main differences with logit models as we know them: first, price is not separated but included in observable characteristics of product j because of the second point, that β_i is a coefficient dependent on consumer characteristics. This means that different consumers will have different tastes in the same characteristics. For example, some person might be interested in the design of a phone while someone would disregard this and focus only on memory, another one on camera quality, etc.

5.1.1 Latent-class models

As a simple case, consider that there are only two types of people, such that only two β_i are possible, say β_H and β_L .

5.1.2 Random coefficients models

Further, we will assume that this "taste variation component" β_i has a nice linear structure such that:

$$\beta_i = \lambda + z_i\beta^o + \nu_i\beta^u$$

thus yielding the following utility:

$$\begin{aligned} U_{ij} &= X_j\lambda + X_jz_i\beta^o + X_j\nu_i\beta^u + \xi_j + \varepsilon_{ij} \\ &= \delta_j + X_jz_i\beta^o + X_j\nu_i\beta^u + \varepsilon_{ij} \end{aligned}$$

where δ_j is the mean utility level as in previous models. In this sense, the linear model we have just described can model logit models if we assume these interaction terms to be 0.

These interactions are of two types (observables and unobservables) and together they remove the IIA problem when aggregating over consumers, since they include utility based on characteristics of people buying the products: this creates a "closeness" between products that was not observed in simple logit models. Nevertheless, at the individual level, the IIA problem remains since for a single consumer, this interaction term disappears and you are left with a simple logit model.

Substitution, Markups and Strategic complementarity

What type of data should you get?

5.2 BLP algorithm

Consider the situation where data is available only on the aggregate product-level, and no consumer data is observable. We will workout the estimation of demand following four steps:

1. Work out the aggregate shares conditional on both δ (mean utility) and β (taste variation).
2. Invert the shares to get ξ .
3. Estimate the model using the method of moments.
4. Repeat until convergence of all parameters.

To help understanding the details of the following steps, you should keep in mind a quick summary of what it to be done. As always, we are interested in the parameters of utility (that affect demand), thus λ and β'' (that we we now call β only). These parameters are estimated using the interaction of the unobservable product characteristic ξ and adequate instruments. Until now, nothing should surprise you as we follow the same steps as described the GMM estimation strategy for logit and nested-logit models. However, this time it is different since to get ξ ,

you will need the parameters you are looking for (this is the main difference of BLP). That is why you use starting values for those, and iterate until you find the parameters that are stable through estimation (this is a fixed point problem).

Step 1: Conditional aggregate shares

Recall that in the simple and nested logit models we studied, the probability P_{ij} that a consumer i buys product j was equal to the market share since they did not differ from the representative consumer in any way. This time, we now that two different consumers will have different probabilities of buying the product. Consequently, the market share is going to be the integral of consumers individual probabilities over their characteristics (here v_i since we do not observe any consumer characteristics).

Our utility function looks like:

$$U_{ij} = \delta_j + X_j v_i \beta^u + \varepsilon_{ij}$$

thus the individual probability of buying product j is given by

$$P_{ij} = \frac{\exp(\delta_j + X_j v_i \beta^u)}{1 + \sum_k \exp(\delta_k + X_k v_i \beta^u)}$$

and finally, integrating this over the space of v , we get the market shares as a function of product characteristics:

$$s_j(\delta, \beta) = \int \frac{\exp(\delta_j + X_j v_i \beta^u)}{1 + \sum_k \exp(\delta_k + X_k v_i \beta^u)} f(v) dv$$

The issue with this integral is that it cannot be solved analytically (v_i is usually a multivariate normal distribution, which makes the market share a multi-dimensional integral); we can approximate it by taking the sample average over a set of ns draws in a simulation. This yields:

$$\hat{s}_j^{ns}(\delta, \beta) = \frac{1}{ns} \sum_i \frac{\exp(\delta_j + X_j v_i \beta^u)}{1 + \sum_k \exp(\delta_k + X_k v_i \beta^u)}$$

which is a function of parameters that we want to estimate (δ and β). We will see that δ will be computed, and β will be estimated by doing this step multiple

times and finding the best one. This is why you should **KEEP YOUR FIRST SIMULATION OVER** ν !! You do not want to change it, else it is never going to converge.

Note that integration over the distribution of ν_i is a different problem than ε_{ij} for which we can use the form of the univariate Gumbel distribution to help us with analysis. In the case of ν_i , we assume a multivariate normal distribution (across multiple consumer unobserved characteristics). If the distribution (not observations, but at least the pdf) is observed, then we can draw from the said distribution.

Moreover, notice that the use of a finite number of draws in the simulation will create a new source of errors within our estimation routine. Enough simulation draws should help tampering this issue, although finding the good number of draws is not an exact science, but more like a tradeoff between computation speed and errors. Overall, numerical evaluation of integrals is a particular topic that is deep enough to think about it carefully.

Step 2: BLP inversion

We now need to recover the product unobservable term ξ_j . In the same way as we did in the simple logit models, we already have the link between δ and ξ , but δ is "buried" in a nonlinear fashion into the the market shares: we need to invert the market shares to get delta, thus ξ , as a function of the shares (rather than the opposite). Doing that requires a special trick that is at the very core of [?] contribution.

Their trick is to use the fact that the following system:

$$\delta_j^k(\beta) = \delta_j^{k-1}(\beta) + \ln(s_j) - \ln(\hat{s}_j^{ns}(\delta^{k-1}, \beta))$$

is a contraction mapping. To see it, understand that s_j is the observed market share, the exponent k represents the iteration process. In other words, by iterating over this function, the δ values will converge to the true value, $\delta^*(\beta, s, \hat{s})$, where s, \hat{s} are respectively the vectors of observed and estimated shares conditional on β . Finally, we can write:

$$\xi(\beta, s, \hat{s}) = \delta^*(\beta, s, \hat{s}) - X_j \lambda$$

and use this form to construct the moments.

Step 3: Constructing the moments

Step 4: Algorithm iteration

The first three steps were performed for a given β

5.3 DFS algorithm

Chapter 6

Product Availability

6.1 Introduction

Previous chapters have introduced demand estimation under

Chapter 7

Entry Models

7.1 Introduction

7.2 Static and exogenous models

The main question of the book from [?] revolves around how do markets evolve to be less or more concentrated? Moreover, Sutton tries to explain why advertising is higher in concentrated industries. In order to explore the answers to these questions, he looks at different market structures and analyzes what makes these structure coherent with the data.

7.2.1 Perfect competition

Consider a market with perfect competition (price-taking assumption), exogenous sunk costs and free entry. The minimum efficient level of production (in the long run) is where $p = \min AC$.

This assumption means that, in the long run, firms will produce only the quantity satisfying this assumption, not more, not less. This implies that the number of firms in a market only depends on the size of the market, M . In particular, if you

denote the quantity produced by firms at the optimum as q^* , you will have that the optimal number of firms n^* is given simply by:

$$n^* = M/q^*$$

As $M \rightarrow \infty$, we also have $n^* \rightarrow \infty$, thus the concentration ratio will tend to 0.

The issue with this simplistic model is that it might not hold in many settings where (1) competition might be imperfect or (2) sunk costs might be endogenous. As an example consider the increase of population in the United States since the creation of Pepsi and Coca-Cola. Even though the population has more than doubled, their respective market shares are still about 30% and 60%, which clearly disproves the previous model. For that reason he explores both directions.

7.2.2 Imperfect competition

Now suppose you allow for imperfect competition. For that purpose, we consider a two-stage game where in the first stage, firms choose to enter a market and incur a cost of A ; then in the second stage, firms play the market game.

Moreover, assume that all M consumers have an income of 1 to spend exclusively on the good produced by the firms, such that in equilibrium:

$$p \cdot Q = M$$

Finally, suppose the marginal cost of production is $c > 0$.

Bertrand competition

Cournot competition

In the Cournot solution, assuming N firms enter, a firm i will choose its quantity q_i such that:

$$q_i \in \arg \max_{q_i} \left[M \cdot \left(\sum_{j=1}^N q_j \right)^{-1} - c \right] \cdot q_i$$

which yields the following FOCs:

$$\left[M \cdot \left(\sum_{j=1}^N q_j \right)^{-1} - c \right] - q_i \cdot M \cdot \left(\sum_{j=1}^N q_j \right)^{-2} = 0 \text{ for all } i = 1, 2, \dots, N$$

$$\Leftrightarrow [M \cdot Q^{-1} - c] - q_i \cdot M \cdot Q^{-2} = 0 \text{ for all } i = 1, 2, \dots, N$$

If we add all N conditions together, we get:

$$N \cdot [M \cdot Q^{-1} - c] - Q \cdot M \cdot Q^{-2} = 0$$

$$\Leftrightarrow N \cdot M \cdot Q^{-1} - N \cdot c - M \cdot Q^{-1} = 0$$

$$\Leftrightarrow \frac{N-1}{N} \cdot M \cdot Q^{-1} = c$$

Finally recall that $M \cdot Q^{-1}$ is actually the equilibrium price, meaning that we have: $p = \frac{N}{N-1}c = (1 + \frac{1}{N-1})c$. Consequently, we have:

$$Q = \frac{N-1}{N} \cdot \frac{M}{c}; q_i = \frac{N-1}{N^2} \cdot \frac{M}{c}; \Pi(N) = \frac{c}{N-1} \cdot \frac{N-1}{N^2} \cdot \frac{M}{c} = \frac{M}{N^2}$$

Hotelling/Salop competition

7.3 Static and endogenous models

7.3.1 Sutton's view

7.3.2 Bresnahan and Reiss (1991)

In a realistic research setting, it is hard to observe strategic variables on a market level (prices, costs, advertising expenses, for all firms). This is the reason why [?] innovated on an estimation using mainly superficial market observables: the number of firms (but not the shares) and other market characteristics. The main assumptions of this model rely on a static Sutton model with symmetric firms and free entry.

Behavioral model

As in the Sutton-type of models, we solve it backwards, assuming the fixed costs of entry are incurred as sunk in the second period, when firms are choosing production.

Demand in the market m is given by:

$$Q_m = d(Z_m, p) \cdot S(Y_m)$$

where $d(\cdot)$ represents the demand function of a representative consumer and $S(\cdot)$ is the total number of consumers that would buy the product. Note that this demand specification has constant returns to scale: doubling S will double Q . Finally, we define the inverse demand curve as $P(Q, Z, Y)$.

Therefore in the second-stage under Cournot competition, each firm solves:

$$\max_{q_i} \Pi_{N,m} \equiv P(q_i, q_{-i}, Z_m, Y_m) \cdot q_i - F_N - C(q_i)$$

where F_N is the endogenous sunk cost associated with N firms entering the market. Without loss of generality (we proved a similar result in the Cournot-Sutton context), assume that in equilibrium, quantities will be symmetric such that $q_i = q_j = q^*$ for all i, j . Because N is already "chosen" in the second stage, we can write $q^* \equiv d(Z_m, p) \cdot \frac{S(Y_m)}{N}$. Then, using this, the profits of each firm is given by:

$$\Pi_{N,m} = P(q^*, q^*, Z_m, Y_m) \cdot q^* - F_N - C(q^*) = \left(P_N - \frac{C(q^*)}{q^*} \right) \cdot d(Z_m, P_N) \cdot \frac{S}{N} - F_N$$

Now, for both the average variable cost $\frac{C(q^*)}{q^*}$ and the fixed cost F_N , let's allow them to be additively separable in components, one of them being dependent on the firm only (respectively AVC and F) and the other component being endogenous on the number of firms (respectively b_N and B_N), such that the total profit of an individual firm, facing a market with N total firms, is defined as:

$$\Pi_{N,m} = (P_N - AVC(q^*) - b_N) \cdot d(Z_m, P_N) \cdot \frac{S}{N} - F - B_N$$

In the first stage, using the free entry assumption, we know that if N^* firms enter a market m , it must be that $\Pi_{N^*,m} > 0$ and $\Pi_{N^*+1,m} < 0$. Conversely, we can look

at the entry threshold s_N , which is the minimum additional demand, not already covered by existing $N - 1$ firms that is required in order for the N th firm to enter. This threshold is the value of $s_N \equiv S_N/N$ such that:

$$\Pi_{N,m} = 0 \Leftrightarrow \frac{S_N}{N} = \frac{F + B_N}{(P_N - AVC_N - b_N)d_N}$$

We can finally also look at the ratio of successive thresholds:

$$s_{N+1}/s_N = \frac{F + B_{N+1}}{F + B_N} \frac{(P_N - AVC_N - b_N)d_N}{(P_{N+1} - AVC_{N+1} - b_{N+1})d_{N+1}}$$

Thus, in a fully competitive market, a firm would enter each time its cost to enter could be recovered in the market, making s_{N+1}/s_N tend to 1, as $N \rightarrow \infty$.

Empirical strategy

The threshold equations derived above ask for a lot of observed variables, namely prices, costs (variable and sunk) and more. In reality, it is very difficult to observe all these at the same time for all firms in a market, thus we need a model that could allow for less information, which is the essence of [?].

Consider a reduced-form profit function given by:

$$\Pi_{N,m} = \underbrace{S(Y_m, \lambda)}_{\text{size}} \cdot \underbrace{V_N(Z_m, W_m, \alpha, \beta)}_{\text{variable profit}} - \underbrace{F_N(W_m, \gamma)}_{\text{fixed}} + \epsilon_m$$

$\bar{\Pi}_{N,m}$

which has an endogenous component $\bar{\Pi}_{N,m}$, and an error-term ϵ_m that is not dependent on N . This should raise some memories to anyone who covered the demand estimation part of the course. In fact, assuming a specification on the error term reveals the model as an ordered probit model, very similar to the logit or multinomial probit models. To see this, consider a market with N firms, this means that:

$$\bar{\Pi}_{N,m} + \epsilon_m > 0 \text{ and } \bar{\Pi}_{N+1,m} + \epsilon_m < 0$$

which is equivalent to:

$$\bar{\Pi}_{N,m} > \epsilon_m > \bar{\Pi}_{N+1,m}$$

Assuming ϵ_m is drawn from an iid normal distribution with mean 0 and variance σ^2 , we have that:

$$\Pr[N_m] = \begin{cases} \Phi(\bar{\Pi}_{N,m}) - \Phi(\bar{\Pi}_{N+1,m}) & \text{if } N_m > 1 \\ 1 - \Phi(\bar{\Pi}_{2,m}) & \text{if } N_m = 1 \end{cases}$$

Before going on to the estimation of the model, we need to look at the reduced-form of each component of the profit function:

- The market size includes a combination of market size characteristics Y_m , such as the population, the neighbouring population, growth, commuting possibilities, etc.
- The variable profit per capita is defined as: $V_N = \alpha_1 + X\beta - \sum_{n=2}^N \alpha_n$ where:
 - X is a vector of relevant economic variables such as demand characteristics for the product (Z) and cost-shifters (W).
 - $\alpha_n \geq 0$ is an intercept component such that each firm entering a market has a negative effect on profits.
- The endogenous fixed costs defined as: $F_N = \gamma_1 + \gamma_L w_L + \sum_{n=2}^N \gamma_n \cdot \gamma_n$

Finally, the estimation is performed using maximum likelihood. In fact, for each observation of the market, we use the probability function defined above. This gives us a likelihood function as a function of all observed variables described in the list above. Taking the log of it yields the log-likelihood to be maximized in order to estimate the parameters of interest.

7.3.3 Berry (1992)

The previous paper was lacking heterogeneity in the sense that its results applies for the case where all firms have the same costs, same continuation payoffs, etc. A step in the direction of allowing some differentiation was made by [?], where fixed costs can vary across firms. Thus, the behavioral model stays identical, but the profit of firm k can now be divided into a common component $v_m(N)$, incurred by all firms in the same way, and an idiosyncratic component $\phi_{m,k}$ applying only to firm k . In particular, we have:

$$\Pi_{N,m,k} = \underbrace{X_m\beta - \delta \ln(N) + \rho u_{m,0}}_{v_m(N)} + \underbrace{Z_k\alpha + \sqrt{1 - \rho^2} \cdot u_{m,k}}_{\phi_{m,k}}$$

Note that in this equation, the combination of the terms $\rho \cdot u_{m,0} + \sqrt{1 - \rho^2} \cdot u_{m,k}$ actually represents the error term, denoted $\epsilon_{m,k}$. In this setting, ρ represents the correlation between error terms across firms in a given market. Finally, [?] assumes that:

$$\epsilon_{m,k} \sim N(0, \Sigma)$$

where Σ is a matrix with all off-diagonal terms equal to ρ .

As in the demand estimation case, this error term is known to the players (the firms) but not to the econometrician. We are still in the static, full information game where all firms know everything about all other firms. However this time we are not in an ordered probit setting since the very structure of the problem is endogenous to the number of firms. To see that, recall the error of the previous model did not include any other subscript than m , while this one includes something about k which is intrinsically linked to the number of firms.

Contrary to the previous models, this one can display multiplicity of equilibria. In fact, although the equilibrium number of firms is unique, the exact firms that enter the market can be different. This implies that we lose information on which firms enter a market. For example, the model might predict that in a given market, two firms will ultimately enter, but you could have firms 1 and 2, firms 2 and 3 or firms 1 and 3. In order to control that issue, the author assumes that firms enter in the order of their profitability. Using that assumption, we can simplify the paper and compute the probability of observing N firms in the market as:

$$\begin{aligned} \Pr[n_m = N | Z_m] &= \Pr \left[\epsilon_{m,1}, \dots, \epsilon_{m,K_m} : \sum_{k=1}^{K_m} \mathbb{I}\{v_m(N, Z_m) > \phi_{m,k}\} = N \right] \\ &= \underbrace{\int \dots \int}_{K_m \text{ times}} \mathbb{I} \left\{ \sum_{k=1}^{K_m} \mathbb{I}\{v_m(N, Z_m) > \phi_{m,k}\} = N \right\} dF(\epsilon_{m,1}, \dots, \epsilon_{m,K_m}, \theta) \end{aligned}$$

which is a multi-dimensional integral (of K_m dimensions) and as we've seen in the previous two chapters, it is hard to compute. In order to circumvent this computational issue, we could use the same method as in BLP: an inner loop simulating the sample moment of the integral for each value of θ , and an outer loop solving for the value of θ that minimizes the square distance between the observed and simulated number of firms. This technique is called the Simulated Nonlinear Least Squares (or SNLS).

Formally, the outer loop finds

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{M} \sum_m (N_m - E [n_m | Z_m, \theta, \epsilon_{m,1}, \dots, \epsilon_{m,K_m}])^2$$

where the expectation term is approximated by its simulated sample average over random iid draws of $(\epsilon_{m,1}, \dots, \epsilon_{m,K_m})$. The simulated sample average is defined as:

$$E [n_m | Z_m, \theta, \epsilon_{m,1}, \dots, \epsilon_{m,K_m}] \approx \frac{1}{S} \sum_s n_m^{*,s}(\epsilon^s, \theta)$$

$$\text{where } n_m^{*,s}(\epsilon^s, \theta) \equiv \max_{0 \leq n \leq K_m} \left\{ n : \sum_{k=1}^{K_m} \mathbb{I}\{v_m(n, Z_m) > \phi_{m,n} | \epsilon^s\} \geq n \right\}$$

or in words, the maximum number of firms n , such that the number of firms entering the market with a positive profit is greater than n .

To recap the full estimation algorithm, consider these steps:

7.4 Two-period models

7.4.1 Stackelberg's model

Two-period models have been introduced for the first time in entry models, but they have been used in several other settings since, for example in models of investment, contracting, etc. Typically, these models rely on:

- A first period where the agents choose a state variable that will determine the nature of the game in the second period.
- A second period where the game is played following the state of the world decided on the first.

The solution concept used to determine the outcome of the game is called "sub-game perfection" which means a Nash equilibrium is chosen at each step played in the game. We also call this equilibrium the Subgame Perfect Nash Equilibrium, or SPNE. In the simple case of a two-period game, the SPNE is quite easy to determine by backwards induction:

- You solve for the NE of the second period game for each potential game played following first period choices.
- Assuming the outcomes computed above are realized, you solve for the first period choice that is a NE.

Two-period capacity game

Consider a typical two-period à la Stackelberg game where a firm chooses a level of capital ($K_1 \geq 0$) in the first period, and the second firm chooses its own level ($K_2 \geq 0$) in the second period. A firm that chooses a capital level of 0 is choosing to stay out of the market. A firm that chooses a level strictly greater than 0 enters the market, although firm 2 would have to pay a fixed cost of entry equal to F .

The profits of firm 1 are:

$$\pi_1(K_1, K_2) = K_1 \cdot (1 - K_1 - K_2)$$

while the profits of firm 2 are:

$$\pi_2(K_1, K_2) = \begin{cases} K_2 \cdot (1 - K_1 - K_2) - F & \text{if } K_2 > 0 \\ 0 & \text{if } K_2 = 0 \end{cases}$$

In order to solve the game, we proceed by backwards induction. In the second period, we consider the decision of firm 2 over the level K_2 . It will choose a level $K_2 > 0$ if and only if the profit associated is greater than 0, meaning that:

$$K_2(1 - K_1 - K_2) - F \geq 0 \Leftrightarrow K_2(1 - K_1 - K_2) \geq F$$

In that case, firm 2 will choose:

$$K_2^* = \arg \max_{K_2} K_2(1 - K_1 - K_2) - F$$

which is the solution to the following FOC:

$$1 - K_1 - 2K_2^* = 0 \Leftrightarrow K_2^* = \frac{1 - K_1}{2}$$

Plugging this solution into the participation condition above, we get that:

$$\frac{1 - K_1}{2} \left(1 - K_1 - \frac{1 - K_1}{2}\right) \geq F \Leftrightarrow \frac{(1 - K_1)^2}{4} \geq F \Leftrightarrow K_1 \leq 1 - 2\sqrt{F}$$

and thus the optimal level K_2^* is defined as:

$$K_2^* = \begin{cases} \frac{1-K_1}{2} & \text{if } K_1 \leq 1 - 2\sqrt{F} \\ 0 & \text{if } K_1 > 1 - 2\sqrt{F} \end{cases}$$

Now that the second period is solved, we can go into the first period. Firm 1 will choose a level K_1 to maximize its profits, which are now known to be:

$$\pi_1(K_1) = \begin{cases} K_1 \cdot \left(1 - K_1 - \frac{1-K_1}{2}\right) & \text{if } K_1 \leq 1 - 2\sqrt{F} \\ K_1 \cdot (1 - K_1) & \text{if } K_1 > 1 - 2\sqrt{F} \end{cases}$$

From this perspective, it is clear that, for a given K_1 , allowing the other player to enter the market is worse than deterring entry. However, deterring entry can only be done for $K_1 > 1 - 2\sqrt{F}$. This creates a discontinuous profit function where deterring entry is always better but can be achieved only at some points that might yield a lower profit than allowing firm 2 to enter. The following graph represents this discrepancy nicely:

7.4.2 Empirical work