Demand Estimation Notes

Frank Pinter*

January 14, 2020

Contents

1 Introduction

> Consumer demand analysis is as old as the sea...
>
> —course description for ECON 2046, which was never taught

Why do we care about demand? Consumer choice theory is not an intrinsic part of the study of industrial organization. In some markets (e.g., government procurement) it does not matter at all, at least as commonly understood. There are three main reasons. The first is that, in most markets, consumer demand provides the incentives for firms to act. Armed with demand estimates, we can understand:

- Pricing decisions
- New product and product repositioning decisions
- Investment decisions: for example, in advertising

because we can write down the payoffs firms get from different courses of action.

The second is that we need to estimate consumer surplus if we want to estimate the welfare effects of a policy, a merger, or a new product. We can't do this without a demand curve.

The third is that, if we are comfortable making assumptions about firm behavior in equilibrium, we can use demand estimates to back out marginal costs without using any direct cost data. A successful example is Nevo's estimates of markups in cereal (Nevo 2001). Obviously we'd rather use cost data if we have it, but you have to start somewhere.

1.1 Discrete choice

Why do we care about discrete choice demand? Sure, at a fundamental level, all choices are discrete: try buying π gallons of milk. But particularly in the types of product markets we work with in IO, discrete choices are a better representation of individual decisions than continuous choices. We didn't run into this in representative agent models: given a large population, the *average* consumer can buy an amount of milk very close to π gallons. That convenience isn't available to us in micro-level models.

That said, in the models considered here, consumers can buy at most one unit of the good. This hasn't often been a problem in IO, but there are markets where it matters.

1.2 Characteristics space

The characteristics space approach supposes that consumers get utility not from goods themselves, but from attributes of those goods. It is philosophically closer to the idea of utility as a psychological object that responds to stimuli. Kelvin Lancaster criticized traditional (product-space) demand theory as follows:

> All *intrinsic* properties of particular goods, those properties that make a diamond quite obviously something different from a loaf of bread, have been omitted from the theory, so that a consumer who consumes diamonds alone is as rational as a consumer who consumes bread alone, but one who sometimes consumes bread, sometimes diamonds (*ceteris paribus*, of course), is irrational. Thus, the only property which the theory can build on is the property shared by all goods, which is simply that they are goods.
>
> —Lancaster (1966)

Even though characteristics space feels like a more fundamental model of differentiated products, as practical researchers, we are happy to use product space models if they suit our purposes better. Plenty of IO papers have done this. Nonetheless, there are practical reasons why characteristics space can work better:

- The too many parameter problem. Even in a simple model, we need to construct J^2 elasticities; the number of parameters that pin these down is on the order of J^2.
- With product space models, we can't make counterfactual predictions if products change or new products are introduced. We can only study new products if we have data post-introduction.

1.3 A note on references

The best reference for traditional discrete choice models is Train's textbook (Train 2009). A good reference for demand modeling in IO is the IO chapter in the Handbook of Econometrics (Ackerberg, Benkard, Berry, and Pakes 2007), which honestly is a good reference for much of modern IO.

These notes were originally written as a study aid for the Harvard Economics graduate field exam in IO. I draw heavily on Ariel Pakes's lecture notes, but at points I take a different perspective. I have also included additional material on the foundational discrete choice models, especially multinomial logit.

2 Historical background

The models we use for discrete choice started with two separate strands within the psychology literature. The first was *random utility maximization* (RUM), which has the following form. Suppose an agent observes multiple alternatives, indexed by j, and each gives the agent utility $V_j + \varepsilon_j$, where V_j is fixed and ε_j is random. What is the probability that the agent chooses a given alternative? Early models derived closed-form expressions for the probability when the ε_j are iid and normally distributed; we now call this the probit model. The early references are Thurstone (1927) and Marschak (1960).

The second was the *Independence from Irrelevant Alternatives axiom* (IIA), which everyone knows is absurd, but you have to start somewhere.[1] Let C be a choice set and let $i, j \in C$. The IIA axiom lets us infer choice probabilities using binomial choices, by assuming that the ratio of any two choice probabilities doesn't depend on the rest of the choice set:

$$\frac{P_C(i)}{P_C(j)} = \frac{P_{\{i,j\}}(i)}{P_{\{i,j\}}(j)}.$$

As we all know today, IIA directly implies the red bus–blue bus problem, first pointed out by Debreu in 1960. Suppose that the only options in $t = 1$ are a train and a red bus, and in $t = 2$ we add a blue bus. If IIA holds and all probabilities are positive, then adding the blue bus reduces the probability of choosing the train. See the appendix for a short proof.

Luce (1959) showed that if all choice probabilities are positive, IIA implies that choice probabilities must have the following form:

$$P_C(i) = \frac{w_i}{\sum_{k \in C} w_k}$$

where the w_i are positive, constant weights that don't depend on the choice set.

A series of papers in the 1960s and 1970s showed the equivalence of IIA and RUM under certain assumptions. In particular, the IIA model is consistent with a RUM model of the form $V_j + \varepsilon_j$ (with ε_j iid) if and only if ε_j is distributed Type 1 Extreme Value, $F(\varepsilon) = \exp(-\exp(-\varepsilon))$.

Early authors used these techniques to model individual choices of travel modes. In the famous BART study, McFadden and coauthors surveyed Oakland and Berkeley residents to collect characteristics of individuals' trips: travel time, waiting time, walking distance, cost, household characteristics, and so on. This was used to estimate a logit model of travel-mode choices for the commute to work. They then did an out-of-sample prediction: what happened when the new BART

[1] Don't confuse this with the various other Independence of Irrelevant Alternatives axioms in economics.

option was added to the choice set? As it turned out, they matched the post-BART market shares closely.

If you are interested in procrastinating, see the Nobel lecture by Dan McFadden (McFadden 2001). All references for this section are available there.

2.1 Differences from representative consumer approach

Think about how someone like Gary Becker would model travel demand. There exists a representative consumer who chooses a continuous allocation across travel modes to maximize a utility function, subject to time and budget constraints. This has some major drawbacks as a framework for empirical work, such as:

- **It tells us nothing about disaggregate data.** The representative consumer framework can't distinguish between a world where everyone takes the bus with some probability, or some proportion of the population always takes the bus. We also can't connect individual choices to individual characteristics.
- **It's a product space model.** We can't use it to evaluate what happens when the goods change, and we can't model new product introduction.
- **It's not econometrically useful.** We don't learn much by taking such a model to the data.

There is a good discussion of this point in McFadden (1981). Modern discrete choice methods can even answer questions like the following: if a new travel mode is introduced, *what kind of person* switches to it? This is beyond the scope of a Becker-style model.

3 The basic multinomial logit model

The most accessible reference is Chapters 2–3 of the textbook by Kenneth Train (Train 2009).

3.1 Random Utility Maximization foundation

Let i index individuals and let j index options. Suppose each option is described by a vector of characteristics $\mathbf{x}_j$, and suppose individual utility is given by a fixed part, which is linear in characteristics, and an unobserved part:

$$u_{ij} = \underbrace{\mathbf{x}_j'\beta}_{\text{fixed}} + \underbrace{\varepsilon_{ij}}_{\text{unobserved}} .$$

The linearity of the utility function is purely for convenience, and it restricts substitution patterns. There is an argument for handling price differently, to allow price sensitivities to vary by income; multiple papers do this, including BLP (Berry, Levinsohn, and Pakes 1995).

Crucially, in the basic discrete choice model, *everyone is the same except for* ε_{ij}. We will only relax this once we get to heterogeneous tastes in section 5.

(We could let the characteristics vary from one person to another. After all, some people live close to the bus stop, and others live far away. In the BART study, x varies from one individual to

Figure 1: The econometrician (U.S. Navy via Wikimedia Commons)

another. In the types of product markets we usually work with in IO, this doesn't normally apply; anyway, we don't usually have fine enough micro data.)

3.2 Choice probabilities

If ε_{ij} is distributed Type 1 Extreme Value across the population, and iid across individuals and products, then *choice probabilities* take the multinomial logit form:

$$s_{ij} \equiv P(j \in \arg\max_{k \in C} u_{ik}) = \frac{\exp(\mathbf{x}_j'\beta)}{\sum_{k \in C} \exp(\mathbf{x}_k'\beta)}.$$

The proof of this is nice. You can find it in Chapter 3 of Train's textbook (Train 2009).

Note that the choice set C must only include choices actually available to individual i. If we have observations from multiple distinct markets, we'll need to take this into account.

3.2.1 Interpretation

For our model to be coherent, the agent must know ε_{ij} when making her decision. The choice is only random *from the econometrician's perspective*. This is why we use unobservables in structural work: we do not see everything that our agents see.

For reasons discussed in section 3.6, the additive error is an unsatisfactory way to handle heterogeneity across the population, and the distributional assumption is made purely for tractability.

3.3 Basic identification issues

Why do we fix the distribution of ε_{ij}? It's a strange distribution: its mean is Euler's constant, $\gamma \approx 0.5772$, and its variance is $\pi^2/6$. Suppose we instead used $\mu + \sigma\varepsilon$:

$$u_{ij} = \mathbf{x}_j'\beta + \mu + \sigma\varepsilon_{ij}.$$

- **Mean utility is not identified.** If we shift utility up by a constant μ for all options, individual choices will never change.

  This means we can normalize the level of utility however we want. Usually we designate an *outside option*, labeled $j = 0$, and normalize $u_{i0} = \varepsilon_{i0}$. The choice of an outside option depends on the particular problem, but we usually think of it as the option whose characteristics are policy-invariant. In product markets, the outside option is not buying the product. Sometimes there is no natural choice of an outside option. (Should we use an outside option when modeling consumers' choice of health insurance? Hospitals?)

  If we are including option-specific dummies in the characteristics vector, we need to fix the coefficient on one option. If we have an outside option, normalizing u_{i0} does this for us.

- **The scale σ is not identified,** which changes the interpretation of our estimates. We can only ever identify the ratio β/σ.

  This means that the estimated β/σ have no independent interpretation — although it's fine to interpret the marginal rates of substitution. (It also means we should be concerned if the scale differs from one population to another. Train discusses this.)

3.4 Estimation

In the basic multinomial logit model, we assume that we can observe all the characteristics, and we only want to estimate β/σ. Our model is fully specified! The only reason why our observed market shares and our model choice probabilities differ is that we have a finite sample.

Since the observed choices are drawn from a multinomial distribution, we can estimate by maximum likelihood. If we have many individuals, and our model is correctly specified, our model's choice probabilities and our observed market shares will match closely.

This differs from the usual approach in IO, which will be introduced in section 4. There we have additional unobservables, whose distributions are only partially specified. This creates other sources of variation we need to account for, even if we have many individuals.

3.5 Expected utility

From the econometrician's perspective, the utility that consumer i gets from her choice is a random variable. This matters for welfare calculations, for example. If a consumer's marginal utility of income α_i is constant over the price region covered by a change, consumer surplus is

$$CS_i = \frac{1}{\alpha_i} \cdot \underbrace{\max_j u_{ij}}_{\text{utility from choice}}$$

Fortunately, in the logit model, there is a convenient closed-form expression for the mean of this random variable:

$$E[\max_j u_{ij}] = \sum_j E[u_{ij} \mid u_{ij} \geq u_{ik} \text{ for all } k] \cdot s_{ij} = \log\left(\sum_j \exp(\mathbf{x}_j'\beta)\right) + c$$

where c is some constant.

3.6 Problems due to unobservables

The basic multinomial logit model is easy to work with because it is so restrictive. It is hard to believe that the characteristics here are the only ones that matter, and misspecification may give us counterintuitive results.

3.6.1 Too many characteristics

If there are too many characteristics relative to the size of our data, we will not get precise estimates of their coefficients. This is just a too many parameters problem. It turns out that we can address this problem by replacing the less-important characteristics with a one-dimensional unobserved characteristic.

3.6.2 Price endogeneity

What if the model is misspecified, and some goods are just better than others in ways the econometrician can't see? Any seller worth its salt will take that into account when setting the price. So in the data, we would see consumers preferring high-priced goods, which would make our estimated price coefficient wrong.

We will address this by explicitly allowing the price of a product to depend on the unobserved characteristic.

4 Adding unobserved quality

The key reference is Berry (1994), whose insight is that adding an unobserved characteristic can help with the endogeneity problem. That paper is also easy to read. To use the Berry model, we need to impose some additional assumptions:

- There must exist a sensible outside good, which we label $j = 0$, with a known market share.
- The market size must be large, so that observed market shares are close to choice probabilities.

Here is the model. Utility is linear in characteristics plus an unobserved quality measure ξ_j for each good, and the logit term:

$$u_{ij} = \underbrace{\mathbf{x}_j'\beta + \xi_j}_{\equiv\delta_j} + \varepsilon_{ij}.$$

We also normalize $u_{i0} = \varepsilon_{i0}$. Note that ξ_j is a vertical characteristic, in the sense that every consumer wants more of it.

4.1 Instrumental variables setup

If ξ_j were exogenous, we could just estimate this model like any other multinomial logit model with alternative-specific dummy variables. But if ξ_j is known to the price-setting process, we would expect it to be correlated with the price p_j. This is referred to as price endogeneity.

Berry's insight is that we can solve this problem if we have instrumental variables. Since economists usually deploy IVs in quasi-experimental settings, this deserves some elaboration. We want to find a variable z_j that affects price directly, while being uncorrelated with ξ_j. In the labor literature, this would be called an "instrument for price". Compare this to a wage regression in labor, where we want an instrument that induces variation in the endogenous x-variable (for example, education) but is not related to unobservable ability.

4.1.1 Moment condition

Suppose that we observe product-level instruments z_j. We will partially specify the distribution of ξ_j, using the conditional moment restriction:

$$E[\xi_j \mid z_j] = 0.$$

(It should be clear from this that if we added a coefficient to ξ_j, it wouldn't be identified.)

To go from a conditional moment restriction to an unconditional moment restriction, we let $h(\cdot)$ be a vector-valued function, and write

$$E[\xi_j h(z_j)] = 0. \tag{1}$$

There are other approaches. For example, with panel data, one could assume that ξ_{jt} follows a first-order Markov process, where innovations are mean-independent of z_j.

4.1.2 Choices of instruments

There are multiple common choices of instruments; the appropriateness of each varies from one application to another. Section 1.4.3 of Ackerberg, Benkard, Berry, and Pakes (2007) has a good discussion of this. In principle, you can use anything that moves prices but is determined after ξ_j is fixed.

- **Exogenous cost shifters.** They're great if you can get them. The idea is that firms can easily respond to cost shifts by changing prices, but not by changing products.
- **Non-price characteristics of the same good.** This is based on a timing assumption: firms first set the observable characteristics, then they observe ξ_j, and then they set prices. This is often justified by saying that prices move more frequently than characteristics do. Armstrong (2016) documents that these instruments become weak if the number of products in a market is reasonably large.
- **Non-price characteristics of other goods.** These are relevant to markups, but don't enter the utility function directly. The most common implementation is "BLP instruments" (Berry, Levinsohn, and Pakes 1995), which use the sum of characteristics of other goods by the same

firm and the sum of characteristics of other goods by other firms:

$$x_{jk}, \quad \sum_{r \neq j, r \in \mathcal{F}_f} x_{rk}, \quad \sum_{r \neq j, r \notin \mathcal{F}_f} x_{rk}.$$

- **Prices in other markets.** These are also known as Hausman instruments; Nevo uses them (Nevo 2001). The idea is that prices elsewhere are a proxy for underlying costs, but are themselves independent of demand shocks in the current market. This is not true if, for example, there has recently been a national ad campaign.

4.2 Estimation

On the face of it, if all we observe are characteristics and market shares, we have a nonlinear IV problem. This is fortunately false. Even though market shares are nonlinear functions of the parameters, we have defined product quality δ_j to be linear in parameters:

$$\delta_j = \mathbf{x}_j'\beta + \xi_j.$$

If we can find a way to back out δ_j from data, we're set: we have a linear IV problem. We can think of the δ_j as a solution to a nonlinear system of J equations in J unknowns:

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_k \exp(\delta_k)} \text{ for all } j. \tag{2}$$

This is where we use the many-consumers assumption, so sampling error in the market shares is small. Berry proves that the system has a unique solution.

4.2.1 The MPEC formulation

The linear IV problem can be written as a constrained optimization problem, following Dubé, Fox, and Su (2012).[2]

Let s_j be the observed market shares. We want the unconditional moment restriction (1) to hold as best as possible, while ensuring that predicted and observed market shares match. First define the sample analog of the unconditional moment condition (1),

$$g(\delta; \beta) = \frac{1}{J} \sum_j (\delta_j - \mathbf{x}_j'\beta) h(z_j).$$

Then solve the GMM problem on these moment conditions, adding the market-share system (2) as constraints:

$$\begin{aligned} \min_{\delta, \beta} \quad & g(\delta; \beta)' W g(\delta; \beta) \\ \text{s.t.} \quad & s_j = \frac{\exp(\delta_j)}{1 + \sum_k \exp(\delta_k)} \text{ for all } j. \end{aligned}$$

[2]This exposition is purely pedagogical. In practice, the Berry inversion is always better for the pure logit model.

Of course, we wouldn't estimate a pure logit model this way, because the Berry inversion provides a simplification.

4.2.2 The Berry inversion

For the logit model, the market share equations have a closed-form solution. This is where we use the outside good. Notice that, for all j,

$$\log \frac{s_j}{s_0} = \log \frac{\exp(\delta_j)}{\exp(0)} = \delta_j.$$

By substituting $\hat{\delta}_j = \log(s_j/s_0)$ for δ, the GMM problem reduces to the following unconstrained problem:

$$\min_{\beta} g(\hat{\delta}; \beta)' W g(\hat{\delta}; \beta)$$

This is equivalent to estimating the following by linear IV:

$$\log \frac{s_j}{s_0} = \mathbf{x}_j' \beta + \xi_j.$$

I will discuss standard errors in section 7. Here, note that the left-hand side has some variance, but we have assumed that sampling error in market shares is small.

4.3 Problems due to IIA

Adding this unobservable does not change the IIA property. IIA becomes particularly vicious when we consider price changes, as a result of the proportionate substitution property: if the share of a product increases, it draws consumers from all other products proportionally. You can see this by looking at the price elasticity formulas. To make the notation easier, break out price from the other characteristics:

$$u_{ij} = \mathbf{x}_j' \beta - \alpha p_j + \xi_j + \varepsilon_{ij}.$$

With some algebra one can show that the own-price derivative of the choice probability is

$$\frac{\partial s_{ij}}{\partial p_j} = -\alpha s_{ij}(1 - s_{ij})$$

while the cross-price derivatives are

$$\frac{\partial s_{ij}}{\partial p_k} = \alpha s_{ij} s_{ik}.$$

This is easy to reject if our data are good enough, and it doesn't line up with our intuition. Ariel Pakes likes the car example: suppose a top-quality expensive Lexus and a terrible cheap Yugo have the same market share. If I raise the price of a high-quality BMW, are the consumers substituting away from the BMW really switching equally to the Lexus and to the Yugo?

To fix this, we need other sources of consumer heterogeneity in addition to ε_{ij}. This will be implemented using random coefficients.

Figure 2: A Yugo (Michael Gil via Wikimedia Commons)

4.3.1 What about nested logit?

The literature first responded to the problems of IIA by relaxing the iid assumption on ε_{ij}. The nested logit model groups products into nests, where IIA holds within a nest but not across nests. (This is similar to the multi-stage budgeting procedure underlying many product space demand systems.) For example, a consumer could decide whether to buy a sedan, SUV, or pickup, then within that choose a quality level, then within that choose a model. You can read more about nested logit in Chapter 4 of Train's textbook (Train 2009), and a version of the Berry (1994) method works for nested logit.

The main difficulty with nested logit is that the nests are arbitrary, and the results depend on the composition and ordering of the nests. (For example, you would get different estimates if consumers choose quality level first, then sedan/SUV/pickup.) Nonetheless, nested logit is still a common demand model in IO, and it might make sense for your application.

5 Adding heterogeneous tastes

We finally arrive at the model of BLP (Berry, Levinsohn, and Pakes 1995), which adds heterogeneity in consumer tastes to the Berry (1994) model. Consumer heterogeneity has three sources:

- **Observed heterogeneity (demographics).** This is used in many applications, particularly Micro BLP (Berry, Levinsohn, and Pakes 2004). It allows us to answer the question: when prices change, or new products are introduced, *who* switches?
- **Unobserved heterogeneity.** This is how we handle heterogeneous tastes that aren't captured by easily measured demographics.
- **Additive ε_{ij}.** Don't forget this is still around, for tractability; we don't generally believe that it helps us capture the data better. Among other things, it ensures that shares are within $(0, 1)$ at any guess of the parameters.

5.1 Model

Somewhat frustratingly, everyone seems to have a different notation, especially for the more complex models. I will attempt to follow the notation of Micro BLP (Berry, Levinsohn, and Pakes 2004) as best as possible; this is almost the same as lecture notes, but not exactly the same.

BLP replaces the constant coefficients with random coefficients.[3] Similarly to before, utility is linear in parameters with an unobservable and an additive error:

$$u_{ij} = \sum_k x_{jk}\beta_{ik} + \xi_j + \varepsilon_{ij}$$

where we replace β with an individual-specific β_i:

$$\beta_{ik} = \overline{\beta}_k + \sum_r z_{ir}\beta^o_{rk} + \nu_{ik}\beta^u_k.$$

Here, z_i is a vector of observable demographics, and ν_i is a vector of unobservables whose distribution is assumed. Typically we assume that the random vector $\nu_{ik}\beta^u_k$ belongs to a parametric family of distributions, such as multivariate normal with a diagonal covariance matrix. Then, assume without loss of generality that $\nu_i \sim N(0, I)$ and subsume the standard deviation in β^u_k. (Nevo (2001) and others allow for correlations between ν_{ik} and $\nu_{i\ell}$, which might be important in applications; we don't consider those here.)

Write out the full utility specification:

$$u_{ij} = \underbrace{\sum_k x_{jk}\overline{\beta}_k + \xi_j}_{\equiv\delta_j} + \sum_k \sum_r x_{jk} z_{ir} \beta^o_{rk} + \sum_k x_{jk}\nu_{ik}\beta^u_k + \varepsilon_{ij}.$$

Exactly as in the logit case, we can extract an overall quality component δ_j that is constant across individuals. Each individual i has the logit choice probability

$$s_j(z_i, \nu_i) = \frac{\exp(\delta_j + \sum_k \sum_r x_{jk} z_{ir} \beta^o_{rk} + \sum_k x_{jk}\nu_{ik}\beta^u_k)}{1 + \sum_q \exp(\delta_q + \sum_k \sum_r x_{qk} z_{ir} \beta^o_{rk} + \sum_k x_{qk}\nu_{ik}\beta^u_k)} \quad \text{for all } j.$$

(To make the notation simpler, I ignore the case where there are multiple distinct markets. In actual applications, always make sure that the logit choice probability is only taken over the products that individual i can access.)

I will return to the demographics in section 8. For now, assume that we don't have any demographic information. By integrating out over the distribution of ν_i in the population, we get market shares:

$$s_j = \int \frac{\exp(\delta_j + \sum_k x_{jk}\nu_{ik}\beta^u_k)}{1 + \sum_q \exp(\delta_q + \sum_k x_{qk}\nu_{ik}\beta^u_k)} \, dF_\nu(\nu_i) \quad \text{for all } j. \tag{3}$$

If β^u is known, this is a nonlinear system of J equations in J unknowns, which has a unique solution. But, in general, β^u is not known. To get it, we need to invoke the IV conditional moment restriction.

[3] Mixed logit models, which have random coefficients but no ξ term, were used well before BLP. See Chapter 6 of Train's textbook (Train 2009).

5.2 The MPEC formulation

One way to do this (MPEC, see Dubé, Fox, and Su (2012)) is by writing the GMM problem and setting (3) as constraints. First define the sample analog of the unconditional moment condition (1),

$$g(\delta;\beta) = \frac{1}{J}\sum_j (\delta_j - \mathbf{x}_j'\beta)h(z_j).$$

Then solve the GMM problem on these moment conditions, adding the market-share system (3) as constraints:

$$\begin{aligned} \min_{\delta,\overline{\beta},\beta^u} \quad & g(\delta;\overline{\beta})'Wg(\delta;\overline{\beta}) \\ \text{s.t.} \quad & s_j = \int \frac{\exp(\delta_j + \sum_k x_{jk}\nu_{ik}\beta_k^u)}{1+\sum_q \exp(\delta_q + \sum_k x_{qk}\nu_{ik}\beta_k^u)}\, dF_\nu(\nu_i) \quad \text{for all } j. \end{aligned}$$

This can be solved directly, and sometimes it's faster to do so. But since we search over δ, the search space grows with the number of products. There is a different way, which takes advantage of the uniqueness of δ given β^u.

5.3 The Berry inversion

The traditional approach instead uses the uniqueness result to create a nested algorithm.

1. Guess a value of β^u
2. Given β^u, solve the nonlinear system of equations to obtain $\hat{\delta}(\beta^u)$
3. Solve the linear IV problem for $\hat{\overline{\beta}}$ and calculate the resulting value of the GMM objective function, $g(\hat{\delta};\hat{\overline{\beta}})'Wg(\hat{\delta};\hat{\overline{\beta}})$
4. Update β^u, and iterate until convergence

This way, we reduce the dimension of the search space. (The effect on runtime is ambiguous, because we now have to solve the system of equations at every guess of the parameters.) To solve the nonlinear system of equations for δ, BLP develop a contraction mapping with a unique solution. Iterate until convergence:

$$\delta_j^{(t+1)} = \delta_j^{(t)} + \log(s_j) - \log\left(\int \frac{\exp(\delta_j^{(t)} + \sum_k x_{jk}\nu_{ik}\beta_k^u)}{1+\sum_q \exp(\delta_q^{(t)} + \sum_k x_{qk}\nu_{ik}\beta_k^u)}\, dF_\nu(\nu_i)\right).$$

Formally, the traditional BLP estimator looks like this:

$$\min_{\overline{\beta},\beta^u} \quad g(\hat{\delta}(s,\beta^u);\overline{\beta})'Wg(\hat{\delta}(s,\beta^u);\overline{\beta})$$

where $\hat{\delta}(s,\beta^u)$ is the limit of the contraction mapping given (s,β^u).

5.4 Computational notes

Considerable ink has been spilled about the best way to implement the BLP estimator, and the problems with the implementations used in the original BLP paper (Berry, Levinsohn, and Pakes 1995) and in early applications, such as Nevo (2001). The upshot is that you can get incorrect results if you're not careful, and there are also tricks that get the estimator to converge faster.

Nowadays we have access to canned methods that avoid implementation pitfalls, such as pyblp in Python and BLPestimatoR in R. If you are implementing the method yourself, take a look at Conlon and Gortmaker (2019) and Brunner, Heiss, Romahn, and Weiser (2017) and the literature cited there.

5.4.1 Calculating integrals

The astute reader will notice that the problems written above can't be taken directly to the data, because the integrals can't be solved exactly. There are many ways to approximate integrals numerically; in this context, simulation-based methods are most common, but quadrature methods are more reliable. In the problem set we used simple Monte Carlo integration, while BLP use a more precise method called importance sampling. Many good textbooks on computational methods cover this material, in addition to Chapter 9 of Train's textbook (Train 2009). For comparisons of integration methods, see Conlon and Gortmaker (2019) and Brunner, Heiss, Romahn, and Weiser (2017).

As section 7 will cover in more detail, using simulation-based approximations to the integral will add variance to the estimate, and needs to be accounted for in the standard errors (as discussed in Berry, Levinsohn, and Pakes (1995) and Berry, Linton, and Pakes (2004).)

5.4.2 GMM details

As GMM problems, the MPEC and Berry formulations both require a choice of the weighting matrix W. For the weighting matrix, Nevo (2001) recommends two-step GMM. First, take the optimal weight matrix under homoskedasticity ($W = (Z'Z)^{-1}$, where Z is the matrix of transformed instruments; see your first-year metrics notes) and run the full procedure. Then recompute $W = (\frac{1}{J}\sum_j Z_j'\xi_j\xi_j'Z_j)^{-1}$ using the estimated values of $\hat{\xi}_j$ from the first step, and run again.

The GMM problems also require a choice of optimization algorithm. As a general principle, use analytic gradients instead of numerical gradients (Dubé, Fox, and Su 2012) and try multiple starting points to avoid getting stuck in local minima. If you are implementing the method yourself, take a look at Conlon and Gortmaker (2019) and Brunner, Heiss, Romahn, and Weiser (2017) and the literature cited there.

6 Adding the supply side

Recall that in homogeneous product markets, we often estimate supply and demand jointly as a system of simultaneous equations, to get more reliable estimates than if we estimated a demand curve alone. The same applies here: adding a pricing model gets us more precise estimates. And for many counterfactuals, we need a model of price setting anyway; we might as well use it here. That said, if we don't have data on cost shifters, we can't do this.

BLP (Berry, Levinsohn, and Pakes 1995) use a marginal cost projection together with a Nash-in-prices equilibrium assumption. Take a linear projection of log marginal cost on a vector of observable cost shifters w_j:

$$\log(mc_j) = w_j\gamma + \omega_j.$$

See Section 1.4.1 of the handbook chapter, or Section 3 of BLP, for the derivation of the pricing equation in a Nash-in-prices equilibrium (or save it for an exercise). If we let $\Delta(p)$ be a $J \times J$ matrix encoding ownership and demand elasticities:

$$\Delta_{jr}(p) = -\frac{\partial s_r}{\partial p_j} \cdot \mathbf{1}[r \text{ and } j \text{ are produced by the same firm}]$$

we obtain (in vector notation)

$$p = mc + \Delta(p)^{-1}\, s(p) \tag{4}$$

where $s(p)$ is the vector of predicted market shares. We can rearrange this to form

$$\omega_j = \log(p_j - e_j'\Delta(p)^{-1}\, s(p)) - w_j\gamma$$

where $e_j'\Delta(p)^{-1}$ is the jth row of $\Delta(p)^{-1}$.

If we plug in our demand estimates, estimating γ is a simple linear IV problem. The insight of BLP is that we can jointly estimate demand parameters ($\overline{\beta}$ and β^u) and γ, using conditional moment restrictions on both ξ_j and ω_j.

BLP make the distributional assumption that we can use the same instruments for ξ_j and ω_j:

$$E[\xi_j \mid z_j] = E[\omega_j \mid z_j] = 0.$$

The suitability of this assumption depends on your use case. Estimation is exactly as above, but the moment function g must be rewritten as

$$g(\delta; \overline{\beta}, \gamma) = \frac{1}{J}\begin{pmatrix} \sum_j (\delta_j - \mathbf{x}_j'\overline{\beta})h(z_j) \\ \sum_j (\log(p_j - e_j'\Delta(p)^{-1}s(p)) - w_j\gamma)h(z_j) \end{pmatrix}$$

and the optimization must be done over γ in addition to the other parameters.

Berry, Levinsohn, and Pakes (1995) report that their demand-only estimates were unreliable with large standard errors, while their joint estimates were sensible. Andrews, Gentzkow, and Shapiro (2017) (a good paper by the way) show on the original BLP data that the estimates are particularly sensitive to violations of the supply-side moment conditions. Nevo (2001), by contrast, doesn't use the supply side in estimation at all. As always, evaluate in the context of your particular problem.

7 Standard errors

Standard errors are discussed in Berry, Levinsohn, and Pakes (1995) and Berry, Linton, and Pakes (2004). The papers show that the asymptotic variance-covariance matrix of the estimates has the usual GMM form. The covariance matrix of the moments is $V_1 + V_2 + V_3$, which capture the three independent sources of variation in our estimates:

1. From the econometrician's perspective, ξ_j is random, and therefore so is the full vector of product characteristics, $(x_j, \xi_j, w_j, \omega_j)$.

   This is exactly the type of error considered in standard GMM asymptotics, so we can handle it with standard GMM methods. Define V_1 to be the standard IV-GMM covariance matrix of the moments that we would use if δ were observed perfectly.

2. We don't actually observe choice probabilities, we observe market shares; there is some sampling error. This is assumed to be negligible in BLP. It could matter in other applications.

3. In aggregating over the population to account for heterogeneity, we introduce sampling error. The form of this error depends on the method used to approximate the integral, and changes if we also sample from demographic data.

   BLP calculate V_3 by a Monte Carlo procedure: at the estimated parameters $(\hat{\beta}^u, \hat{\overline{\beta}})$, draw a new set of ν_i, recalculate the integral, and recalculate the empirical moments. (They find that this matters.)

8 Adding additional data

There are multiple ways to add demographic and micro data, and the right way depends on the data you have. Recall the market share equation when we have observable consumer characteristics:

$$s_j = \iint \frac{\exp(\delta_j + \sum_k \sum_r x_{jk} z_{ir} \beta^o_{rk} + \sum_k x_{jk} \nu_{ik} \beta^u_k)}{1 + \sum_q \exp(\delta_q + \sum_k \sum_r x_{qk} z_{ir} \beta^o_{rk} + \sum_k x_{qk} \nu_{ik} \beta^u_k)} \, dF_z(z_i) \, dF_\nu(\nu_i) \quad \text{for all } j.$$

8.1 Using market-level distributions

Suppose that we have the distributions of demographic variables in each market. For example, the Census gives us the joint distribution of income and household size at the regional level. We can just integrate over F_z the same way we integrate over F_ν. The procedure is exactly the same, but rather than searching over β^u in the GMM problem, we search over (β^u, β^o).

One way to do this is to simultaneously draw $z_i \sim F_z, \nu_i \sim F_\nu$ and calculate the integral by simulation. Nevo's cereal study (Nevo 2001) does this for data on income, age, and number of children.

8.2 Additional moment restrictions

Petrin's study of the minivan (Petrin 2002) adds moment restrictions from the Consumer Expenditure Survey, which provides correlations between consumer demographics and the products they purchase. Though ill-equipped to work as micro data, this still helps pin down the demand model. Petrin adds moment conditions that match CEX averages with model predictions. If you were curious, the particular moments are:

- Probability of purchasing a new vehicle, given income (discretized into three buckets)
- Average family size given the type of vehicle purchased
- Probability the head of household is 30–60 years old, given the type of vehicle purchased

Formally, these model predictions take the form:

$$g_2(\delta;\theta) = f(\delta,\theta) - \mathbf{m}$$

where $f(\delta,\theta)$ is the predicted vector of moments given parameters (e.g., $\theta = (\overline{\beta}, \beta^o, \beta^u, \gamma)$), and $\mathbf{m}$ is the vector of actual moments from CEX. For estimation, just stack these moments $g_2(\cdot)$ with the IV moments $g(\cdot)$ and run BLP as usual. For standard errors, it helps that variation in the CEX moments is independent of the variation in the rest of the process, so the covariance matrix of the moments is block-diagonal.

This method helps to estimate the effects of interactions between consumer observables and product characteristics (the β^o). This helps us model substitution better, and report the ways substitution patterns relate to demographics. It is only mildly helpful with β^u.

8.3 Micro BLP

Micro BLP (Berry, Levinsohn, and Pakes 2004) obtain one year of the CAMIP survey, which connects consumer demographics to the products they purchase and, crucially, consumers' first choices to their second choices. Why are second choices useful? The unobservable tastes $\beta^u\nu_i$ are important determinants of substitution patterns. Second choices provide (hypothetical) data on substitution patterns: they tell us what a consumer would do if her choice set were changed. From this, we can back out information about unobservable tastes.

Micro BLP focuses on estimating $(\delta, \beta^o, \beta^u)$, which it does by matching moments. After this is complete, they try a few ways to estimate $\overline{\beta}$ given δ.

8.3.1 Estimation: first step

The moments they use from the CAMIP survey are:

- Covariances of first-choice product characteristics $\mathbf{x}$ and consumer demographics z_i
- Covariances of first-choice product characteristics and second-choice product characteristics

Since the CAMIP survey is choice-based, they need to include a correction when calculating the covariance.

For intuition, the problem can be written in MPEC form as:

$$\begin{aligned}
\min_{\delta,\beta^o,\beta^u} \quad & m(\delta;\beta^o,\beta^u)' W m(\delta;\beta^o,\beta^u) \\
\text{s.t.} \quad & s_j = \iint \frac{\exp(\delta_j + \sum_k \sum_r x_{jk} z_{ir} \beta^o_{rk} + \sum_k x_{jk}\nu_{ik}\beta^u_k)}{1 + \sum_q \exp(\delta_q + \sum_k \sum_r x_{qk} z_{ir} \beta^o_{rk} + \sum_k x_{qk}\nu_{ik}\beta^u_k)} \, dF_z(z_i)\, dF_\nu(\nu_i) \text{ for all } j
\end{aligned}$$

where $m(\delta;\beta^o,\beta^u)$ is the difference between model-implied moments at $(\delta;\beta^o,\beta^u)$ and the moments from CAMIP.

For the actual implementation, Micro BLP applies the contraction mapping from BLP to solve the market-share inversion and obtain $\hat{\delta}(\beta^o,\beta^u)$. Then the problem becomes

$$\min_{\beta^o,\beta^u} \quad m(\hat{\delta}(\beta^o,\beta^u);\beta^o,\beta^u)' W m(\hat{\delta}(\beta^o,\beta^u);\beta^o,\beta^u)$$

which is solved to obtain $(\hat{\beta}^o, \hat{\beta}^u)$ and $\hat{\delta} = \hat{\delta}(\hat{\beta}^o, \hat{\beta}^u)$.

8.3.2 Estimation: second step

Equipped with estimates $(\hat{\delta}, \hat{\beta}^o, \hat{\beta}^u)$, the last step is to break down $\hat{\delta}$ into an estimate of $\overline{\beta}$. Recall that

$$\delta_j = \sum_k x_{jk}\overline{\beta}_k + \xi_j.$$

Micro BLP tries three different methods, none of which works especially well. The challenge is that we only have one cross-section of data, so there isn't much variation to work with. The three methods are:

1. Set the price coefficient to zero
2. Joint supply-and-demand estimation with linear IV, like in BLP
3. Calibrate the market-level price elasticity to one (suggested by staff at GM)

9 Remaining problems due to logit term

We still have the ε_{ij} random term in utility, and it can cause us some trouble. To get an intuition for the role of the logit term in driving substitution patterns, consider the extreme cases:

- **Pure logit model.** All consumer heterogeneity comes from the ε_{ij} term; everyone is equally likely to substitute to the new good.
- **Logit term is negligible.** Each consumer has an 'ideal product' and chooses the product that is closest to it. Only consumers whose current purchases are very similar to the new product would be willing to switch.

In the intermediate case, the random term makes every good at least a little bit desirable to everyone: every good has a nonzero choice probability for every consumer. This means that if we add a terrible product, we will overstate its new market share as the random term drives some people to it. If we add a superior product, we will understate its new market share as the random term continues to drive some people to their old, worse options. Likewise, when the price of a product rises, we will overstate the number of consumers that stick with it.

If these problems are especially severe for your application, Ariel Pakes (Ackerberg, Benkard, Berry, and Pakes 2007) would suggest using the pure characteristics model instead, which drops the ε_{ij} terms. (If you can make it computationally tractable for your problem, that is.)

10 Examples

10.1 Thought experiments

Train your intuition about substitution with heterogeneous agents.

1. Suppose there are two products in the market, made by different firms. Firm 1 is suddenly hit with a cost shock and has to raise its prices. Should firm 2 raise its prices in response?[4]

   Not necessarily. Who substitutes away from product 1? The most price-sensitive people, of course. Firm 2 may find it profitable to lower its prices and steer these people away from the outside good.

2. Here's the example from lecture. We observe in the data that after the oil price shocks of the early 1970s, the average fuel efficiency of new cars got worse. How does this make sense?

   Who drops out of the car market? Poorer people, who now have less money to spend on cars because gas is expensive, and who also tend to buy small, fuel-efficient cars. That leaves only richer people, who tend to buy larger, less fuel-efficient cars.

   As Ariel Pakes likes to point out, BLP could explain this well, but it could not explain the improvements in fuel efficiency that occurred a few years later. We would need a dynamic model to explain the carmakers' decision to respond to oil price shocks by developing different cars.

10.2 Calculating markups

How can we estimate markups if we don't have direct cost estimates, or even cost shifters? With an equilibrium assumption, we might be able to back out markups from demand parameters and market outcomes. An example of this is Nevo (2001), who combines demand-only BLP estimation with the Nash-in-prices assumption under a few different ownership scenarios. Recall that the pricing equation (4) has a strong implication for markups:

$$p - mc = \Delta(p)^{-1} s(p)$$

where $s(p)$ are shares, and $\Delta(p)$ is a function of ownership and elasticities, which in turn depend only on the primitives we estimated.

10.3 Merger simulation

The effect of a merger depends on the equilibrium assumption. For Nash-in-prices, recall that the pricing equation (4) is a fixed point of:

$$p = mc + \Delta(p)^{-1} s(p).$$

Recall that $\Delta(p)$ encodes ownership information. By changing the ownership data in the matrix $\Delta(p)$, we can simulate the new prices by calculating the fixed point of the pricing equation.

10.4 New product introduction (ex ante)

Suppose we have data on a market, and we want to predict what would happen if a new product were introduced.

[4] That is, are prices strategic complements or strategic substitutes?

- **We need to know its characteristics.** In particular, we need a model of its unobservable characteristic ξ_j. Micro BLP (Berry, Levinsohn, and Pakes 2004) construct the predicted ξ_j from the estimated ξ_j of other products from the same manufacturer.
- **We need to know its price.** There is no right way to do this; Micro BLP use the prediction from a regression of price on product characteristics and manufacturer dummies.
- **We need to know the responses of competitors.** Micro BLP explicitly shut this down. If we instead assume competitors can respond by changing prices, we could use a pricing equation from a particular equilibrium model, like Nash-in-prices.

(Of course, these are not a problem if we get to see pre- and post-data, as Petrin (2002) does: all of these objects can be observed or estimated. In particular, many of our problems are solved if we allow ξ_j to change for all goods in the post-period.)

Because of the problems discussed in section 9, our estimates for new product introduction will be biased toward the pure logit results, which suffer from the red bus–blue bus problem. In particular, if we add a bad product, we will overstate its new market share, and if we add a good product, we will understate its market share.

A Red bus–blue bus proof

For notation, let p_1 denote choice probabilities in the first period over $\{T, R\}$, and let p_2 denote choice probabilities in the second period over $\{T, R, B\}$. Let $\alpha = p_1(R)/p_1(T)$. In the first period, then,

$$1 = p_1(T) + p_1(R) = (1+\alpha)p_1(T) \implies p_1(T) = \frac{1}{1+\alpha}.$$

In the second period, IIA implies that $\alpha = p_2(R)/p_2(T)$ as well. Denote

$$r = \frac{p_2(B)}{p_2(R)}.$$

Suppose that the blue bus gets some market share, so that $r > 0$. It follows that

$$1 = p_2(T) + p_2(R) + p_2(B) = p_2(T)(1+\alpha+\alpha r) \implies p_2(T) = \frac{1}{1+\alpha+\alpha r} < \frac{1}{1+\alpha} = p_1(T).$$

References

Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes (2007). "Econometric Tools for Analyzing Market Outcomes". In: *Handbook of Econometrics*. Vol. 6. Elsevier, pp. 4171–4276. ISBN: 978-0-444-50631-3. DOI: 10.1016/S1573-4412(07)06063-1. URL: http://linkinghub.elsevier.com/retrieve/pii/S1573441207060631.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro (2017). "Measuring the Sensitivity of Parameter Estimates to Estimation Moments". In: *The Quarterly Journal of Economics* 132.4, pp. 1553–1592. DOI: 10.1093/qje/qjx023.

Armstrong, Timothy B. (2016). "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply". In: *Econometrica* 84.5, pp. 1961–1980. ISSN: 1468-0262. DOI: 10.3982/ECTA10600.

Berry, Steven (1994). "Estimating Discrete-Choice Models of Product Differentiation". In: *The RAND Journal of Economics* 25.2, pp. 242–262. DOI: 10.2307/2555829. URL: https://www.jstor.org/stable/2555829.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile Prices in Market Equilibrium". In: *Econometrica* 63.4, pp. 841–890. DOI: 10.2307/2171802.

— (2004). "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market". In: *Journal of Political Economy* 112.1, pp. 68–105. DOI: 10.1086/379939.

Berry, Steven, Oliver B. Linton, and Ariel Pakes (2004). "Limit theorems for estimating the parameters of differentiated product demand systems". In: *The Review of Economic Studies* 71.3, pp. 613–654. DOI: 10.1111/j.1467-937x.2004.00298.x. URL: http://restud.oxfordjournals.org/content/71/3/613.short.

Brunner, Daniel, Florian Heiss, André Romahn, and Constantin Weiser (2017). *Reliable estimation of random coefficient logit demand models*. DICE Discussion Paper 267. URL: http://hdl.handle.net/10419/168359.

Conlon, Christopher and Jeff Gortmaker (2019). *Best Practices for Differentiated Products Demand Estimation with pyblp*. Working paper. URL: https://chrisconlon.github.io/site/pyblp.pdf.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su (2012). "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation". In: *Econometrica* 80.5, pp. 2231–2267. DOI: 10.3982/ECTA8585.

Lancaster, Kelvin J. (1966). "A New Approach to Consumer Theory". In: *Journal of Political Economy* 74.2, pp. 132–157. DOI: 10.1086/259131. URL: http://www.jstor.org/stable/1828835.

McFadden, Daniel (1981). "Econometric Models of Probabilistic Choice". In: *Structural Analysis of Discrete Data with Econometric Applications*. Ed. by Charles F. Manski and Daniel McFadden. Cambridge, MA: MIT Press, pp. 198–272. URL: https://eml.berkeley.edu/~mcfadden/discrete/ch5.pdf (visited on 05/06/2018).

— (2001). "Economic choices". In: *The American Economic Review* 91.3, pp. 351–378. DOI: 10.1257/aer.91.3.351. URL: http://www.jstor.org/stable/2677869.

Nevo, Aviv (2001). "Measuring Market Power in the Ready-to-Eat Cereal Industry". In: *Econometrica* 69.2, pp. 307–342. DOI: 10.1111/1468-0262.00194. URL: http://www.jstor.org/stable/2692234.

Petrin, Amil (2002). "Quantifying the Benefits of New Products: The Case of the Minivan". In: *Journal of Political Economy* 110.4, pp. 705–729. DOI: 10.1086/340779.

Train, Kenneth (2009). *Discrete choice methods with simulation*. 2nd. Cambridge University Press. DOI: 10.1017/cbo9780511805271. URL: https://eml.berkeley.edu/books/choice2.html (visited on 05/06/2018).