Demand Estimation

Paolo Zacchia

Microeconometrics

Lecture 14

Demand estimation: overview

- Demand functions are a key component of economic theory and model. How to **estimate** them empirically?

- This discussion starts by reviewing **traditional** approaches to the econometric estimation of demand systems.

- More modern approaches to demand estimation are based on limited dependent variable **random utility models**, which are grounded on models of **multinomial choice**.

- This helps build up tools and concepts useful for introducing the **workhorse** econometric model by Berry, Levinsohn and Pakes (1995, **BLP**) for market-level product data.

- Lastly, this lecture overviews relevant extensions of the BLP framework and the research frontier on the topic.

Traditional approaches to demand estimation

- Traditional models for demand estimation are based firmly on **microeconomic theory** through explicit specifications of the utility or the cost/expenditure functions.

- These models were developed with the objective of providing decent **approximations** to the true – **unknown** – demand systems, using flexible functional form specifications.

- The very first model in this class, perhaps, was the so-called **Linear Expenditure System** (LES).

- The two main models that are sometimes still used nowadays are the **"translog"** demand system and the **"Almost Ideal Demand System"** (AIDS).

- Before proceeding, it is useful to characterize some common notation and review some results of microeconomic theory.

Shared notation

In what follows, given a **consumer** indexed by i:

- $\boldsymbol{y}_i = (Y_{1i}, Y_{2i}, \ldots, Y_{Ji})$ are J **products** consumed by i;
- $\boldsymbol{p} = (P_1, P_2, \ldots, P_J)$ are their corresponding J **prices**;
- M_i is the **individual income** of consumer i;
- $\boldsymbol{s}_i = (S_{1i}, S_{2i}, \ldots, S_{Ji})$ are J **budget shares** defined as:

$$S_{ji} = \frac{P_j Y_{ji}}{M_i} \quad \text{for } j = 1, \ldots, J;$$

- $U(\boldsymbol{y}_i, M_i)$ is the **direct** utility function for income M_i;
- $V(\boldsymbol{p}, M_i)$ is the **indirect** utility function for income M_i;
- $C(\boldsymbol{p}, U_i)$ is the **expenditure/cost** function for utility U_i.

Logarithmic Shepard's Lemma

- Recall Shepard's Lemma from microeconomic theory:

$$\frac{\partial C\left(\boldsymbol{p}, U_i\right)}{\partial P_j} = Y_{ji}^c\left(\boldsymbol{p}_i, U_i\right)$$

  where $Y_{ji}^c\left(\boldsymbol{p}, U_i\right)$ is the **Hicksian** (compensated) demand for product j.

- Observe that, if one works with **logarithms**:

$$\frac{\partial \log C\left(\boldsymbol{p}, U_i\right)}{\partial \log P_j} = \frac{P_j}{C\left(\boldsymbol{p}, U_i\right)} \frac{\partial C\left(\boldsymbol{p}, U_i\right)}{\partial P_j} = S_{ji}\left(\boldsymbol{p}, U_i\right)$$

  which is an expression for the **budget share** $S_{ji}\left(\boldsymbol{p}, U_i\right)$ of product j as a function of prices $\boldsymbol{p}$ and utility U_i.

- One can work out an analogous result for Roy's identity and Marshallian, rather than Hicksian, demand.

Logarithmic Roy's Identity

- Also recall Roy's Identity from microeconomic theory:

$$-\left(\frac{\partial V\left(\boldsymbol{p},M_i\right)}{\partial P_j}\right)\left(\frac{\partial V\left(\boldsymbol{p},M_i\right)}{\partial M_i}\right)^{-1}=Y_{ji}\left(\boldsymbol{p}_i,M_i\right)$$

  where now $Y_{ji}\left(\boldsymbol{p},U_i\right)$ is the **Marshallian** (uncompensated) demand for product j.

- Similarly as before, if one works with **logarithms**:

$$-\frac{\dfrac{\partial\log V\left(\boldsymbol{p},M_i\right)}{\partial\log P_j}}{\dfrac{\partial\log V\left(\boldsymbol{p},M_i\right)}{\partial\log M_i}}=-\frac{\dfrac{P_j}{V\left(\boldsymbol{p},M_i\right)}\dfrac{\partial V\left(\boldsymbol{p},M_i\right)}{\partial P_j}}{\dfrac{M_i}{V\left(\boldsymbol{p},M_i\right)}\dfrac{\partial V\left(\boldsymbol{p},M_i\right)}{\partial M_j}}=S_{ji}\left(\boldsymbol{p},M_i\right)$$

  which is an expression for the **budget share** $S_{ji}\left(\boldsymbol{p},M_i\right)$ of product j as a function of prices $\boldsymbol{p}$ and income M_i.

The elasticities of interest

The estimation of a demand system ideally allows to recover:

- the **Marshallian price elasticities**, for $\ell = 1, \ldots, J$:

$$\eta_{Y_{ji}}^{P_\ell} \equiv \frac{P_\ell}{Y_{ji}(\boldsymbol{p}, M_i)} \frac{\partial Y_{ji}(\boldsymbol{p}, M_i)}{\partial P_\ell} = \frac{\partial \log Y_{ji}(\boldsymbol{p}, M_i)}{\partial \log P_\ell}$$

- …the **income elasticity**:

$$\eta_{Y_{ji}}^{M} \equiv \frac{M_i}{Y_{ji}(\boldsymbol{p}, M_i)} \frac{\partial Y_{ji}(\boldsymbol{p}, M_i)}{\partial M_i} = \frac{\partial \log Y_{ji}(\boldsymbol{p}, M_i)}{\partial \log M_i}$$

- …and the **Hicksian price elasticities**, for $\ell = 1, \ldots, J$:

$$\eta_{Y_{ji}^c}^{P_\ell} \equiv \frac{P_\ell}{Y_{ji}^c(\boldsymbol{p}, U_i)} \frac{\partial Y_{ji}^c(\boldsymbol{p}, U_i)}{\partial P_\ell} = \frac{\partial \log Y_{ji}^c(\boldsymbol{p}, U_i)}{\partial \log P_\ell}$$

- …all related via the **Slutsky** equation $\eta_{Y_{ji}}^{P_\ell} = \eta_{Y_{ji}^c}^{P_\ell} - \eta_{Y_{ji}}^{M} M_i$.

Some general considerations

- All the models that follow are system of equations where the **endogenous** variables are either $\boldsymbol{y}_i$, $\boldsymbol{s}_i$ or $M_i\boldsymbol{s}_i$. The prices $\boldsymbol{p}$ are typically treated as **exogenous**.

- These models can be estimated on either **individual-level** or **"aggregate"** (e.g. market-level) data, depending on the available level of variation in the key variables.

- The individual microeconomic foundations may not hold on average in the population (**"aggregation problem"**), but they *do hold* for the AIDS, which indeed is "almost ideal."

- Several extensions of the models presented hereinafter exist. Typically, their objective is to make the models more general and robust. Only the baseline models are reviewed here.

Linear Expenditure System (1/2)

- The **Linear Expenditure System** (LES) is most famously associated with Geary (1954) and Stone (1955).

- The LES was originally conceived to make sense of household expenditure patterns at a time of scant data availability.

- Assume the following utility function:

$$U\left(\boldsymbol{y}_i; M_i\right) = \prod_{j=1}^{J} \left\{ \left(Y_{ji} - \mu_j\right)^{\beta_j} \cdot \mathbb{1}\left[Y_{ji} > \mu_j\right] \right\}$$

  where μ_j is the **subsistence level** for product j.

- The Marshallian demand for product $j = 1, \ldots, J$ is derived as follows.

$$Y_{ji} = \mu_j + \frac{\beta_j}{P_j} \left(M_i - \sum_{k=1}^{J} P_k \mu_k \right)$$

Linear Expenditure System (2/2)

- This Marshallian demand yields an **econometric model**:

$$P_j Y_{ji} = \alpha_j + \beta_j M_i + \varepsilon_{ji}$$

where ε_{ji} is an **additive** consumer-specific error, and:

$$\alpha_j \equiv \left(P_j \mu_j - \sum_{k=1}^{J} P_k \mu_k \right)$$

is a product-specific constant. This model can be estimated with household-level data about income M_i and expenditure by product (category) $P_j Y_{ji}$.

- This model is interesting because the parameter β_j allows to calculate the income elasticity of demand, which is equal to β_j / S_{ji} for consumer/household i.

- However, the model is simplistic and plagued by endogeneity.

Translog demand system (1/3)

- "Translog" stands for **"trascendental logarithmic:"** this model was originally introduced by Christensen, Jorgenson, and Lau (1975).

- The starting point is a specification of indirect utility.

$$\log V\left(\boldsymbol{p}, M_i\right) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \log\left(\frac{P_j}{M_i}\right) + \\ + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk} \log\left(\frac{P_j}{M_i}\right) \log\left(\frac{P_k}{M_i}\right)$$

  As usual, this is invariant to monotonic transformations.

- The **"trascendental"** part is the summation of logarithmic cross-products, which is meant to approximate higher-order curvatures of the *true* indirect utility function.

Translog demand system (2/3)

- Applying the logarithmic Roy's identity here gives:

$$S_{ji} = \frac{\alpha_j + \sum_{k=1}^{J} \beta_{jk} \log\left(\frac{P_k}{M_i}\right)}{\overline{\alpha} + \sum_{k=1}^{J} \overline{\beta}_k \log\left(\frac{P_k}{M_i}\right)}$$

where $\overline{\alpha} = \sum_{j=1}^{J} \alpha_j$ and $\overline{\beta}_k \equiv \sum_{j=1}^{J} \beta_{jk}$ for $k = 1, \ldots, J$.

- For $\ell = 1, \ldots, J$, the Marshallian price elasticity is:

$$\eta_{Y_{ji}}^{P_\ell} = -\mathbb{1}\left[j = \ell\right] + \frac{\beta_{j\ell}/S_{ji} - \sum_{k=1}^{J} \beta_{jk}}{\overline{\alpha} + \sum_{k=1}^{J} \overline{\beta}_k \log\left(\frac{P_k}{M_i}\right)}$$

- ...whereas the income elasticity of demand is as follows.

$$\eta_{Y_{ji}}^{M} = 1 + \frac{-\sum_{k=1}^{J} \beta_{jk}/S_{ji} - \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}}{\overline{\alpha} + \sum_{k=1}^{J} \overline{\beta}_k \log\left(\frac{P_k}{M_i}\right)}$$

Translog demand system (3/3)

- To empirically estimate the model, econometricians typically specify an **additive** error term ε_{ji} so that the model can be estimated by NLLS with household-level or aggregate data.

$$S_{ji} = \frac{\alpha_j + \sum_{k=1}^{J} \beta_{jk} \log\left(\frac{P_k}{M_i}\right)}{\overline{\alpha} + \sum_{k=1}^{J} \overline{\beta}_k \log\left(\frac{P_k}{M_i}\right)} + \varepsilon_{ji}$$

- Note: as $\sum_{j=1}^{J} S_{ji} = 1$ for all $i = 1, \ldots, N$, this means that one out of J error terms is residually determined!

- Therefore, this is a model of $J-1$ equations with $J\,(J-1)$ right-hand side variables *à la* $\log\left(P_k/M_i\right)$ as well as $J\,(J+1)$ parameters *à la* α_j and β_{ij}. This calls for **restrictions**.

- Theory delivers the **normalization** $\overline{\alpha} = -1$, the **symmetry** property $\beta_{jk} = \beta_{kj}$, and **homogeneity**: $\overline{\beta}_k = \sigma\alpha_k$.

Almost ideal demand system (1/4)

- The **"Almost Ideal Demand System"** (AIDS), which is leading among the traditional approaches, is associated with the seminal contribution by Deaton and Muellbauer (1980).

- The starting point is a specification of the cost/expenditure function for a **representative consumer** with $U_i \in [0, 1]$.

$$\log C(\boldsymbol{p}, U_i) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \log(P_j) + \\ + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{*} \log(P_j) \log(P_k) + U_i \gamma_0 \prod_{j=1}^{J} P_j^{\gamma_j}$$

  This follows from **aggregation-invariant** preferences.

- **Homogeneity** of this cost function demands **restrictions**: $\sum_{j=1}^{J} \alpha_j = 1$, and $\sum_{k=1}^{J} \beta_{jk}^{*} = \sum_{j=1}^{J} \beta_{jk}^{*} = \sum_{j=1}^{J} \gamma_j = 1$.

Almost ideal demand system (2/4)

- Applying the logarithmic Shepard's lemma here gives:

$$S_{ji} = \alpha_j + \sum_{k=1}^{J} \beta_{jk} \log(P_k) + U_i \gamma_i \gamma_0 \prod_{j=1}^{J} P_j^{\gamma_j}$$

  where $\beta_{jk} = \frac{1}{2}\left(\beta^*_{jk} + \beta^*_{kj}\right)$.

- Writing **total expenditures** $X_i = C(\boldsymbol{p}, U_i)$, solving for U_i, and substituting gives:

$$S_{ji} = \alpha_j + \sum_{k=1}^{J} \beta_{jk} \log(P_k) + \gamma_i \log\left(\frac{X_i}{P}\right)$$

  where P is a **price index** defined as follows.

$$\log(P) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \log(P_j) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \log(P_j) \log(P_k)$$

Almost ideal demand system (3/4)

- Observe the similarity of $\log(P)$ with $\log V(\boldsymbol{p}, M_i)$ from the translog model: Lewbel (1989) noted that both the translog and the AIDS models can be nested into a more general one.

- The translog parts within the *demand functions themselves* let interpret the AIDS as a good **approximation** of the *true* system of demand functions (hence its name).

- For $\ell = 1, \ldots, J$, the Marshallian price elasticity is:

$$\eta_{Y_{ji}}^{P_\ell} = -\mathbb{1}[j = \ell] + \frac{\beta_{j\ell} - \gamma_j \left(\alpha_\ell + \sum_{k=1}^{J} \beta_{\ell k} \log(P_k)\right)}{S_{ji}}$$

- ... whereas the income elasticity of demand is as follows.

$$\eta_{Y_{ji}}^{M} = \frac{\gamma_j}{S_{ji}} + 1$$

Almost ideal demand system (4/4)

- Like in the translog model, estimation requires the inclusion of an additive error term ε_{ji}; one out of J obtains residually.

$$S_{ji} = \alpha_j + \sum_{k=1}^{J} \beta_{jk} \log(P_k) + \gamma_i \log\left(\frac{X_i}{P}\right) + \varepsilon_{ji}$$

- This model should be estimated via NLLS if P is explicitly made a function of all the parameters; in most applications however using an **external price index** is preferred, as this makes the system one of $J-1$ (simpler) **linear** equations.

- Identification entails considerations similar to those from the translog case: theory-based **restrictions** are thus necessary.

- Hence, the restrictions that ensure **homogeneity** of the cost function and the **symmetry** property $\beta_{jk} = \beta_{kj}$ are upheld.

Issues with the traditional approaches (1/2)

- The imposition of theoretical Slutsky "curvature conditions" require even more, possibly complex, restrictions.

- With time series, autocorrelation in the errors complicates the specification of the identifying restrictions.

- An important problem is the **curse of dimensionality**: the number of parameters grows *quadratically* with the number of products, which exacerbates any possible statistical issues.

- The estimation of so many cross-product price elasticities for possibly unrelated products can quickly become too unstable and therefore not credible.

- An implication of this curse is the **"new good problem:"** specifically, researchers are unable to analyze the impact of a new product prior to its introduction.

Issues with the traditional approaches (2/2)

- In general all these models were developed with great care for the underlying theory, but with less regard for **endogeneity**. The error terms are likely to include omitted variables!

- More generally, **supply is absent** from traditional models. One notable attempt to incorporate supply is the model by Bresnahan (1987), which inspired subsequent treatments of the supply side in studies about market power.

- Traditional models do not account for similarity between two products' **observable characteristics**. Yet, cross-product price elasticities arguably depend upon product similarities.

- Perhaps most importantly among all issues, **heterogeneity across consumers** is totally ignored. This is likely to **bias** the results whether based on individual or aggregate data.

Beyond traditional approaches

All these limitations stimulated methodological research that led to the current modern takes on demand estimation.

A brief intellectual history is thus sketched.

1. It starts with the venerated model by Bresnahan (1987), to detect collusion (a **supply** mechanism) in oligopolies.

2. It follows through with the analysis of the **random utility** framework by Berry (1994).

3. It culminates with the full-fledged econometric treatment of Berry's original framework: the one by Berry, Levinsohn and Pakes (1995), the BLP model, which is founded on a **mixed logit** that is nested in a larger GMM model.

4. It then concludes with the important extension of BLP by Nevo (2001), which focuses on statistical **identification**.

Detecting collusion in oligopolies (1/7)

- Industrial Organization is currently the "structural" field of economics *par excellance*.
- Decades ago however it saw little empirical work. Classical questions, like the one regarding the sudden 45% increase in the US automobile production and sales in 1955 (that some attributed to a temporary price war), were left unanswered.
- Paul Samuelson himself had allegedly once said that he… *would flunk any econometrics paper that claimed to provide an explanation of 1955 auto sales.*
- Yet Bresnahan (1987) defied Samuelson as he developed an original model estimated via MLE that can be used to *test for collusion.* The results suggest that a price war occurred.
- Although this model is now obsolete, it is still a classic and has great instruction value.

Detecting collusion in oligopolies (2/7)

- Consider N types of cars each with **quality** $X_i = X(\boldsymbol{z}_i, \boldsymbol{\beta})$ being a function of one car's characteristics $\boldsymbol{z}_i$ (given some parameters $\boldsymbol{\beta}$). Qualities can be ordered from best to worst: without loss of generality, $X_i > X_h$ if $i > h$.

- Also consider some well-microfounded **demand functions** of each car for each year $t = 1, \ldots, T$:

$$Q_{it}^D = D(P_{ht}, P_{it}, P_{jt}, X_{ht}, X_{it}, X_{jt}, \boldsymbol{\gamma})$$

where Q_{it} is the **quantity** of product i, P_{it} its **price**, h, i, j are three **consecutive** products in the **order** of "qualities" (with $j > i > h$) and $\boldsymbol{\gamma}$ are some parameters.

- This specification makes prices and quantities dependent in equilibrium only on those of the "neighbors" of one product in the product space.

Detecting collusion in oligopolies (3/7)

It is useful to elaborate on the particular demand function used by Bresnahan. He assumes a simple linear **utility** for consumers:

$$U(X, P, v) = vX - (Y - P)$$

where Y is consumers' income and $v \sim \mathcal{U}(0, V_{max})$ their **taste**. The total "mass" of the consumer base is given by Δ.

Given the previous example about product ordering, a **marginal consumer** indifferent between i and h at time t has taste:

$$v_{hit} = \frac{P_{it} - P_{ht}}{X_{it} - X_{ht}}$$

from which one obtains the demand function as follows.

$$Q_{it}^{D} = \Delta \left[\frac{P_{jt} - P_{it}}{X_{jt} - X_{it}} - \frac{P_{it} - P_{ht}}{X_{it} - X_{ht}} \right]$$

This is based on Prescott and Visscher (1977) as well as Shaked and Sutton (1983).

Detecting collusion in oligopolies (4/7)

Supply is standard: the profits from the sale of product i are:

$$\pi_{it} = P_{it}Q_{it} - c\left(X_{it}\right)Q_{it}$$

with $c\left(X_{it}\right) = \mu \exp\left(X_{it}\right)$. Bresnahan analyzes two scenarios.

1. **Competition**: the firm setting the price of product i takes that of "neighbors" h and j as given, hence this FOC for P_{it}.

$$\frac{\partial \pi_{it}}{\partial P_{it}} = Q_{it} + \left(P_{it} - c\left(X_{it}\right)\right)\frac{\partial Q_{it}\left(\cdot\right)}{\partial P_{it}} = 0$$

2. **Cooperation**: the firm(s) selling two products i and j set prices P_{it} and P_{jt} to maximize their joint profits, with FOCs for price P_{it} as follows (and symmetrically for P_{jt}).

$$\frac{\partial\left[\pi_{it} + \pi_{jt}\right]}{\partial P_{it}} = Q_{it} + \left(P_{it} - c\left(X_{it}\right)\right)\frac{\partial Q_{it}\left(\cdot\right)}{\partial P_{it}} + \left(P_{jt} - c\left(X_{jt}\right)\right)\frac{\partial Q_{jt}\left(\cdot\right)}{\partial P_{it}} = 0$$

Detecting collusion in oligopolies (5/7)

- Bresnahan then defines several **matrices $\mathbf{H}_t$** such that, in each year $t = 1, \ldots, T$:

$$h_{(ij)t} = \begin{cases} 1 & \text{cooperation between products } i \text{ and } j \\ 0 & \text{competition between products } i \text{ and } j \end{cases}$$

  and he characterizes several hypothetical scenarios for 1955 and surrounding years with corresponding matrices $\mathbf{H}_t$.

- Thus, for a **given choice** of matrix $\mathbf{H}_t$ the supply function can be written as

$$q_{it}^S = S\left(P_{ht}, P_{it}, P_{jt}, X_{ht}, X_{it}, X_{jt}, \mathbf{H}_t, \boldsymbol{\gamma}, \mu\right)$$

  where the demand parameters $\boldsymbol{\gamma}$ enter via the derivative of the demand functions implied in the supply FOCs.

- This sets up alternative **counterfactuals** about collusion.

Detecting collusion in oligopolies (6/7)

- By setting the equilibrium condition $Q_{it}^{D} = Q_{it}^{S} = Q_{it}^{*}$ (and similarly for prices) for each product $i = 1, \dots, N$ in every year $t = 1, \dots, T$, the **reduced form** is derived.

$$P_{it} = P^{*}\left(X_{ht}, X_{it}, X_{jt}, \mathbf{H}_t, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mu\right)$$
$$Q_{it} = Q^{*}\left(X_{ht}, X_{it}, X_{jt}, \mathbf{H}_t, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mu\right)$$

- This requires to develop the full-fledged solution given the assumptions about demand and supply, and a matrix $\mathbf{H}_t$.

- Introduce some error terms that make the model stochastic, and endow them of distributional assumptions as follows.

$$\begin{pmatrix} P_{it} - P^{*} \\ Q_{it} - Q^{*} \end{pmatrix} = \begin{pmatrix} \xi_{it}^{P} \\ \xi_{it}^{Q} \end{pmatrix} = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_P^2 & 0 \\ 0 & \sigma_Q^2 \end{pmatrix}\right)$$

- All these error terms are implicit functions of the **reduced form** “predictions” of the model.

Detecting collusion in oligopolies (7/7)

- For any **given** matrix $\mathbf{H}_t$, the model is estimated via MLE using the following likelihood function.

$$\mathcal{L}\left(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mu \,\middle|\, \left\{\mathbf{H}_t, \{p_{it}, q_{it}, \mathbf{z}_{it}\}_{i=1}^{N}\right\}_{t=1}^{T}\right) =$$

$$= \prod_{t=1}^{\mathrm{T}} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_P^2}} \exp\left(-\frac{\left(\xi_{it}^{P}\right)^2}{2\sigma_P^2}\right) \frac{1}{\sqrt{2\pi\sigma_Q^2}} \exp\left(-\frac{\left(\xi_{it}^{Q}\right)^2}{2\sigma_Q^2}\right)$$

- Tests about alternative matrices $\mathbf{H}_{t0}$ and $\mathbf{H}_{t1}$ (say, one for "competition" and one for "collusion") are performed via the likelihood ratio approach, comparing alternative estimates.

$$\mathrm{C}_{\mathbf{H}} = 2\left[\log \mathcal{L}\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \widetilde{\mu} \,\middle|\, \mathbf{H}_{t1}, \dots\right) - \log \mathcal{L}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\mu} \,\middle|\, \mathbf{H}_{t0}, \dots\right)\right]$$

- This way, Bresnahan was able to reject the hypothesis that the the 1955 sales obtained from a competitive, rather than collusive scenario (and vice versa).

Towards the workhorse BLP framework

- Bresnahan's paper was ingenious and was met with success, but its treatment of the demand side is very restrictive.

- This stimulated work on extensions of the multinomial logit (baseline, nested, mixed, etc.) to provide a more flexible and realistic treatment of the demand side.

- Notice that multinomial models are directly applicable when researchers have access to the individual **microdata** about Y_i (for example, consumers' individual purchases).

- How to make a good use of them in aggregate **market-level** data, when researchers have access to the **prices** P_j, market shares S_j and characteristics $\boldsymbol{x}_j$ of individual products j?

- The final objective is to calculate **cross-price elasticities** while overcoming the problems of traditional approaches to demand estimation.

Why random utility models?

Random utility models: extended limited dependent variable models with multinomial responses, are the backbone of modern approaches to demand estimation for several reasons.

- They are grounded upon **latent variable** representations of utility that is dependent on **product characteristics**.
- Thus, the number of parameters scales with the **number of characteristics** – not with the number of products.
- They naturally allow for individual heterogeneity at the cost of using **simulation-based** estimation approaches.
- They naturally allow estimation on both **individual-level** as well as **aggregate** data.

Random utility models however do not solve **endogeneity** issues, which must be appropriately accounted for.

Product choice with random utilities (1/8)

- All starts with the following **random utility** representation of consumers' preferences for one of J competing goods, as in Berry (1994) and subsequent contributions.

$$V_{ji} = \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{\beta}_i - \alpha P_j + \xi_j + \varepsilon_{ji}$$

- This expression features **random coefficients $\boldsymbol{\beta}_i$** for some K product characteristics $\boldsymbol{x}_j$, and **random shock** ξ_j that is product-specific: this is the unobserved "average quality" of product j, net of (random) individual evaluations ε_{ji}.

- With micro-data, this model could be numerically estimated via mixed logit, provided that adequate measures are taken to deal with ξ_j and possibly, the **endogeneity** of prices with respect to ε_{ji}.

- Typically though, only market-level data are available.

Product choice with random utilities (2/8)

Assume that, for $k = 1, \dots, K$, the random coefficients are:

$$\beta_{ki} = \beta_k + \sigma_k \upsilon_{ki}$$

where $\sigma_k \geq 0$ is a parameter and υ_{ki} is an error term specific to the i-th consumer. This is like a mixed logit model where matrix $\mathbf{\Sigma}$ is restricted to its diagonal (tastes are uncorrelated across $\boldsymbol{x}_j$).

Thus, the model can be rewritten as:

$$V_{ji} = \delta_j + \nu_{ji}$$

where:

$$\delta_j \equiv \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{\beta} - \alpha P_j + \xi_j$$

is the **mean utility** of product j, with $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$; while:

$$\nu_{ji} \equiv \sum_{k=1}^{K} X_{kj} \sigma_k \upsilon_{ki} + \varepsilon_{ji}$$

is the mean-zero, heteroscedastic **random component** of utility.

Product choice with random utilities (3/8)

- To make the model more realistic for use with market-level data, assume the existence of an **"outside good"** $j = 0$.
- This can be thought of as the "outside option" of not buying any of the competing J products.
- The mean utility of the outside good is **normalized** at zero.

$$\delta_0 = 0$$

- Consumers still have random preferences $\nu_{01} = V_{0i}$ over it.
- Ultimately, consumers' choice is determined as:

$$Y_{ji} = 1 \;\Leftrightarrow\; V_{ji} = \max\{V_{0i}, V_{1i}, \ldots, V_{Ji}\}$$

for $j = 0, 1, \ldots, J$ (the choice set includes the outside good).

Product choice with random utilities (4/8)

Given a stochastic structure for ν_{ji}, the **market shares** S_j that are predicted by the model are, for $k = 0, 1, \ldots, J$:

$$S_j\left(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\sigma}\right) = \int_{\mathbb{R}^K} \prod_{k \neq j} \mathbb{1}\left[\delta_j + \nu_{ji} \geq \delta_k + \nu_{ki}\right] f_{\boldsymbol{\nu}}\left(\boldsymbol{\nu}, \boldsymbol{X}; \boldsymbol{\sigma}\right) \mathrm{d}\boldsymbol{\nu}$$

where:

- $\boldsymbol{\delta} = (\delta_0, \delta_1, \ldots, \delta_J)$ are the products' mean utilities;
- $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_K)$ are the loadings of the random coefficients;
- $\boldsymbol{\nu} = (\nu_{0i}, \nu_{1i}, \ldots, \nu_{Ji})$ are consumer i's random utilities;
- $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_J \end{bmatrix}$ is the random matrix of characteristics.

This provides a full specification of the **demand system** in terms of market shares (share of consumers who buy product j) S_j.

Product choice with random utilities (5/8)

The key idea by Berry (1994) is to **invert** the demand system so as to solve for $\boldsymbol{\delta}$, and thus estimate the key parameters through the **linear** model:

$$\delta_j\left(\boldsymbol{s}\right) = \boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta} - \alpha P_j + \xi_j$$

where δ_j is implicitly expressed as a function of the market shares $\boldsymbol{s} = (S_0, S_1, \ldots, S_J)$, and where ξ_i is treated as an error term.

- This makes the model estimable on **market-level** data!
- This model may be estimated via OLS, however IV-2SLS is perhaps preferable due to endogeneity concerns.
- The solution $\delta_j\left(\boldsymbol{s}\right)$ embeds all consumers' optimal choice.
- The key question is whether $\boldsymbol{s}$ can be **uniquely** solved for $\boldsymbol{\delta}$. Leveraging Brower's fixed point theorem, Berry shows that a unique solution **always** exists.

Product choice with random utilities (6/8)

For example, if $\boldsymbol{\sigma} = \mathbf{0}$ and $\varepsilon_{ji} \sim \text{Gumbel}\left(0, 1\right)$ for $j = 0, 1, \dots, J$ and all individuals i, the model's realization probabilities (market shares) have the multinomial logit structure:

$$S_j\left(\boldsymbol{\delta}\right) = \frac{\exp\left(\delta_j\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_k\right)}$$

because $\exp\left(\delta_0\right) = 1$. As this also implies:

$$\frac{S_j}{S_0} = \exp\left(\delta_j\right)$$

the solution for mean utilities is straightforward in this case.

$$\delta_j = \log\left(S_j\right) - \log\left(S_0\right) = \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{\beta} - \alpha P_j + \xi_j$$

While analytically convenient, this version of the model inherits all the limitations of the multinomial logit, like homogeneity and IIA (with implications in terms of price elasticities).

Product choice with random utilities (7/8)

Given a solution $\boldsymbol{\delta}(\boldsymbol{s})$, how to address **endogeneity** in the mean utility model? Berry (1994) treats the product characteristics $\boldsymbol{x}_j$ as exogenous, but he also allows for prices to be correlated with unobserved product quality: $\mathbb{E}[P_j \xi_j] \neq 0$.

To address this, he suggests to leverage the **supply side** through appropriate **cost shifters**, like in the canonical model of demand and supply in partial equilibrium.

To this end, it is useful to write the First Order Conditions from the maximization of firm profits in terms of share elasticities:

$$P_j = C(\boldsymbol{w}_j, \omega_j; \boldsymbol{\gamma}) + \frac{S_j}{|\partial S_j(P_j) / \partial P_j|}$$

where $C(\boldsymbol{w}_j, \omega_j; \boldsymbol{\gamma})$ is a **marginal cost function**, $\boldsymbol{w}_j$ is a set of cost shifters with parameters $\boldsymbol{\gamma}$, and ω_j an unobserved error.

Product choice with random utilities (8/8)

Suppose *for example* that the marginal cost function is **linear**.

$$C\left(\boldsymbol{w}_j, \omega_j; \boldsymbol{\gamma}\right) = \boldsymbol{w}_j^{\mathrm{T}} \boldsymbol{\gamma} + \omega_j$$

The objective is to **jointly estimate** a **simultaneous system of equations** that deliver identification of α. Observe that here:

$$\left| \frac{\partial S_j\left(P_j\right)}{\partial P_j} \right| = \alpha \frac{\partial S_j\left(\delta_j\right)}{\partial \delta_j}$$

and therefore the system in this case would be:

$$\begin{aligned} \delta_j\left(\boldsymbol{s}\right) &= \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{\beta} - \alpha P_j + \xi_j \\ P_j &= \boldsymbol{w}_j^{\mathrm{T}} \boldsymbol{\gamma} + \frac{1}{\alpha} \frac{S_j}{\partial S_j\left(\delta_j\right) / \partial \delta_j} + \omega_j \end{aligned}$$

which is typically relatively easy to handle once $\partial S_j\left(\delta_j\right) / \partial \delta_j$ is derived from the choice probabilities. For example, in the simple multinomial logit case it is $\partial S_j\left(\delta_j\right) / \partial \delta_j = S_j\left(1 - S_j\right)$.

The workhorse BLP framework (1/9)

- The idea of "inverting" market shares S_j for mean utilities δ_j is smart, but at first it seemed hard to implement.

- Besides the simple, yet problematic multinomial logit case, Berry (1994) showed that the model allows for **closed form** solutions if the consumers' choice problem has a *nested logit* structure, or if the products are *vertically differentiated* like in Shaked and Sutton (1983) and Bresnahan (1987).

- Yet this is not enough: a full solution to the general **random coefficients** structure of the choice problem is necessary.

- This requires **simulations**, but it was not clear at first how to embed them in the estimation framework.

- The importance of the contribution by Berry, Levinsohn and Pakes (1995, BLP) lies in their solution for this problem.

The workhorse BLP framework (2/9)

- Given $(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\theta})$, with $\varepsilon_{ji} \sim \text{Gumbel}(0,1)$ for $j = 0, 1, \ldots, J$, and given a set of **simulated** K-long vectors $\{\boldsymbol{\upsilon}_s\}_{s=1}^{S}$, it is:

$$\widehat{S}_j(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\sigma}) = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp\left(\delta_j + \boldsymbol{x}_j^{\mathrm{T}} (\mathbf{I}\boldsymbol{\sigma}) \boldsymbol{\upsilon}_s\right)}{\sum_{k=1}^{J} \exp\left(\delta_k + \boldsymbol{x}_k^{\mathrm{T}} (\mathbf{I}\boldsymbol{\sigma}) \boldsymbol{\upsilon}_s\right)}$$

for $j = 1, \ldots, J$, and where $\mathbb{V}\text{ar}[\boldsymbol{u}_s] = \mathbf{I}$ for $s = 1, \ldots, S$.

- Let $\widehat{\boldsymbol{s}}(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\sigma}) = \left(\widehat{S}_1(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\sigma}), \ldots, \widehat{S}_J(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\sigma})\right)$.

- Also let $\boldsymbol{s} = (S_1, \ldots, S_J)$ be the **actual** market shares.

- Berry, Levinsohn and Pakes (1995) show that the **operator** $T(\boldsymbol{\delta}; \boldsymbol{\sigma}) : \mathbb{R}^J \rightarrow \mathbb{R}^J$ defined as:

$$T(\boldsymbol{\delta}; \boldsymbol{\sigma}) = \boldsymbol{\delta} + \log(\boldsymbol{s}) - \log(\widehat{\boldsymbol{s}}(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\sigma}))$$

is a **contraction** for $\boldsymbol{\delta}$ with modulus that is less than one.

The workhorse BLP framework (3/9)

- Introduce the index $m = 1, \ldots, M$ to denote markets as the **unit of observation**.

- Express $\boldsymbol{z}_{jm} = (Z_{1jm}, \ldots, Z_{Qjm})$ as the **instruments** vector of length Q that is specific to product j and market m.

- Write the following **moment conditions**:

$$\mathbb{E}\left[\boldsymbol{z}_{jm}\begin{pmatrix} \widehat{\delta}_{jm}\left(\boldsymbol{s}_m; \boldsymbol{\sigma}\right) - \boldsymbol{x}_{jm}^{\mathrm{T}}\boldsymbol{\beta} - \alpha P_{jm} \\ P_{jm} - \boldsymbol{w}_{jm}^{\mathrm{T}}\boldsymbol{\gamma} - \dfrac{1}{\alpha}\dfrac{S_{jm}}{\partial S_{jm}\left(\delta_{jm}\right)/\partial \delta_{jm}} \end{pmatrix}\right] = \mathbf{0}$$

  also written more compactly as $\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{q}_{jm}; \boldsymbol{\theta}\right)\right] = \mathbf{0}$, where:

$$\begin{aligned} \boldsymbol{q}_{jm} &= \left(P_{jm}, \boldsymbol{s}_m, \boldsymbol{w}_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{z}_{jm}\right), \\ \boldsymbol{\theta}_1 &= \left(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}\right), \\ \boldsymbol{\theta} &= \left(\boldsymbol{\theta}_1, \boldsymbol{\sigma}\right). \end{aligned}$$

- Note that identification requires that at least $2Q \geq |\boldsymbol{\theta}|$.

The workhorse BLP framework (4/9)

The **BLP estimator** is a GMM estimator.

$$\widehat{\boldsymbol{\theta}}_{BLP} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\min} \left[\overline{\mathbf{g}}_{M,J}(\boldsymbol{\theta})\right]^{\mathrm{T}} \mathbf{A}_{2Q} \left[\overline{\mathbf{g}}_{M,J}(\boldsymbol{\theta})\right]$$

with $\overline{\mathbf{g}}_{M,J}(\boldsymbol{\theta}) \equiv \frac{1}{MJ} \sum_{m=1}^{M} \sum_{j=1}^{J} \boldsymbol{g}(\boldsymbol{q}_{jm}; \boldsymbol{\theta})$ and weights $\mathbf{A}_{2Q}$.

This estimator is calculated via a famous **nested fixed point**, "inner loop, outer loop" algorithm. At every iteration of $\boldsymbol{\theta}$:

- in the **inner loop**, $\widehat{\boldsymbol{\delta}}_m$ is calculated *for each market* m as:

  $$\boldsymbol{\delta}_m^{h+1} = \boldsymbol{\delta}_m^h + \log(\mathbf{s}_m) - \log\left(\widehat{\boldsymbol{s}}\left(\boldsymbol{\delta}_m^h, \mathbf{X}_m; \boldsymbol{\sigma}\right)\right)$$

  and iterating over $h = 0, 1, 2, \ldots$ until convergence;

- in the **outer loop**, the particular value of $\boldsymbol{\theta}$ that minimizes the GMM objective function is sought.

The workhorse BLP framework (5/9)

It is useful to recapitulate the "BLP algorithm." Starting with a "guess" of $\boldsymbol{\sigma}$, iterate the following four sub-steps:

1. for $m = 1, \ldots, M$, **solve** for $\widehat{\boldsymbol{\delta}}_m \left(\boldsymbol{s}_m, \mathbf{X}_m; \boldsymbol{\sigma}\right)$ using the inner loop contraction proposed by BLP;

2. for $m = 1, \ldots, M$ and $j = 1, \ldots, J$, **evaluate** the elasticity components $S_{jm} \left[\partial S_{jm} \left(\delta_{jm}\right) / \partial \delta_{jm}\right]^{-1}$ at $\widehat{\boldsymbol{\delta}}_m$;

3. **aggregate** the market-level data and, **given** the elasticities thus evaluated, derive an estimate of $\boldsymbol{\theta}_1$ *implied by* $\boldsymbol{\sigma}$ using IV-2SLS on the now **linear** demand and supply equations;

4. given such implicit estimates $\boldsymbol{\theta}_1 \left(\boldsymbol{\sigma}\right)$, **compute** the *empirical moments* $\overline{\mathbf{g}}_{M,J} \left(\boldsymbol{\theta}_1 \left(\boldsymbol{\sigma}\right), \boldsymbol{\sigma}\right)$, hence the GMM objective;

until convergence at some $\widehat{\boldsymbol{\sigma}}_{BLP}$ and $\widehat{\boldsymbol{\theta}}_{1,BLP} = \boldsymbol{\theta}_1 \left(\widehat{\boldsymbol{\sigma}}_{BLP}\right)$.

The workhorse BLP framework (6/9)

Some observations are due.

- Optimizing over $\boldsymbol{\theta}_1$ given $\boldsymbol{\sigma}$ is simple (linear algebra suffices). The numerical problem is to search for the optimal $\boldsymbol{\sigma}$ in the outer loop. Also, the inner loop may slow down the search.

- According to a recent review (Berry and Haile, 2021): *"while many authors succeeded in implementing and customizing the BLP algorithm, naïve implementations can easily fail."*

- The choice of the GMM weighting matrix $\mathbf{A}_{2Q}$ and statistical inference are both guided by the standard theory of GMM.

- The model allows for more general **marginal cost** functions $C(\boldsymbol{w}_j, \omega_j; \boldsymbol{\gamma})$. In their original formulation, BLP even allow for more general firm First Order Conditions, that obtain if firms "control" multiple products (like in Bresnahan, 1987).

The workhorse BLP framework (7/9)

How to choose the instruments vector $\boldsymbol{z}_{im}$?

- Clearly, the price P_{jm} as well as the market share S_{jm} must be **excluded** from it (they are endogenous).

- The standard "BLP instruments" are based on the **product characteristics** of **other**, potentially substitute, products. The key idea is that substitutability affects markups/prices and market shares. An instrument set for X_{kjm} may include:

$$\boldsymbol{z}_{(kjm)} = \left(X_{kjm} \qquad \sum_{\ell \neq j, \ell \in \mathcal{J}_{jm}} X_{k\ell m} \qquad \sum_{\ell \neq \mathcal{J}_{jm}} X_{k\ell m} \right)$$

  where $\mathcal{J}_{jm}$ is the set of products by the firm that produces product j in market m.

- A similar argument applies to **cost shifters**, like the costs for materials or energy, taxes or tariffs.

- Other instruments have been proposed in the literature.

The workhorse BLP framework (8/9)

- Recall that the main objective of demand estimation is the calculation of price elasticities. How is this performed in the BLP framework? Here, treat θ as known.

- Own-price elasticities are expressed in given a market m as:

$$\eta_{S_{jm}}^{P_{jm}} = -\alpha \frac{P_{jm}}{S_{jm}} \int_{\mathbb{R}^K} p_{jm}(\boldsymbol{\nu}, \boldsymbol{X}) \left[1 - p_{jm}(\boldsymbol{\nu}, \boldsymbol{X})\right] f_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{X}) \, \mathrm{d}\boldsymbol{\nu}$$

- …while cross-price elasticities are as follows, for $\ell \neq j$.

$$\eta_{S_{jm}}^{P_{\ell m}} = \alpha \frac{P_{jm}}{S_{jm}} \int_{\mathbb{R}^K} p_{jm}(\boldsymbol{\nu}, \boldsymbol{X}) \, p_{\ell m}(\boldsymbol{\nu}, \boldsymbol{X}) \, f_{\boldsymbol{\nu}}(\boldsymbol{\nu}, \boldsymbol{X}) \, \mathrm{d}\boldsymbol{\nu}$$

- Here $p_{\ell m}(\boldsymbol{\nu}, \boldsymbol{X})$ is the logit probability for $\ell = 1, \dots, J$.

$$p_{\ell m}(\boldsymbol{\nu}, \boldsymbol{X}) \equiv \frac{\exp\left(\delta_{\ell m} + \boldsymbol{x}_{\ell m}^{\mathrm{T}} (\mathbf{I}\boldsymbol{\sigma}) \boldsymbol{\nu}\right)}{\sum_{k=1}^{J} \exp\left(\delta_{km} + \boldsymbol{x}_{km}^{\mathrm{T}} (\mathbf{I}\boldsymbol{\sigma}) \boldsymbol{\nu}\right)}$$

The workhorse BLP framework (9/9)

- BLP tested their framework for the first time on data about the US market for cars from 1971 to 1990.

- Their definition of "market" is thus a year, which led them to take measures to address autocorrelation of ξ_{jm} and ω_{jm}.

- Their estimates $\widehat{\boldsymbol{\theta}}_{BLP}$ and their simulation draws $\{\boldsymbol{v}_s\}_{s=1}^{S}$ are used to simulate own-price and cross-price elasticities in each market m; the results are then averaged out.

- Estimates of a restricted model without random coefficients ($\boldsymbol{\sigma} = \mathbf{0}$) return implausible, too inelastic demand functions for many products. The extended BLP model delivers more realistic estimates, especially in very competitive segments.

- Random coefficients also dramatically improve realism of the price elasticities to the *outside good* (that is, buying no car).

The improved BLP framework (1/5)

- The current standard for estimation of the BLP model was set by Nevo (2001) with his study on the ready-to-eat cereal industry. He also allows for **random coefficients** for **price**.

$$V_{ji} = \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{\beta}_i - \alpha_i P_j + \xi_j + \varepsilon_{ji}$$

- The random coefficients $(\alpha_i, \boldsymbol{\beta}_i)$ are now specified as:

$$\begin{pmatrix} \alpha_i \\ \boldsymbol{\beta}_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\Pi} \boldsymbol{d}_i + \boldsymbol{\Sigma} \boldsymbol{\upsilon}_i$$

  where $\boldsymbol{\upsilon}_i = (\upsilon_{Pi}, \upsilon_{1i}, \dots, \upsilon_{Ki})$ like in BLP (with the addition of υ_{Pi}), $\boldsymbol{d}_i = (d_{1i}, \dots, d_{Di})$ is a vector of D **demographic characteristics** typical of consumer i's market, while $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma}$ are two matrices of parameters of dimension $(K+1) \times D$ and $(K+1) \times (K+1)$ respectively.

- This leads to a more realistic treatment of consumers' choice: preferences vary as a function of a market's demographics.

The improved BLP framework (2/5)

- Write $\theta = (\theta_1, \theta_2)$, where θ_1 is as before while $\theta_2 = (\mathbf{\Pi}, \mathbf{\Sigma})$. Usually $\mathbf{\Sigma}$ is restricted to the diagonal like in BLP: $\mathbf{\Sigma} = \mathbf{I}\boldsymbol{\sigma}$.

- Also write the random vector $\boldsymbol{r}_j = (P_j, \boldsymbol{x}_j)$ of length $K + 1$. The market share of product j, for $j = 1, \dots, J$, obtains as:

$$\widehat{S}_j \left(\boldsymbol{\delta}, \boldsymbol{p}, \boldsymbol{X}; \theta_2\right) = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp\left(\delta_j + \boldsymbol{r}_j^{\mathrm{T}} \left[\mathbf{\Pi}\mathbf{d}_s + \mathbf{\Sigma}\boldsymbol{\upsilon}_s\right]\right)}{\sum_{k=1}^{J} \exp\left(\delta_k + \boldsymbol{r}_k^{\mathrm{T}} \left[\mathbf{\Pi}\mathbf{d}_s + \mathbf{\Sigma}\boldsymbol{\upsilon}_s\right]\right)}$$

  where $\boldsymbol{p} = (P_1, \dots, P_J)$; $\{\boldsymbol{\upsilon}_s\}_{s=1}^{S}$ is a vector of **simulation draws** as before, now each of length $K + 1$...

- ... whereas $\{\mathbf{d}_s\}_{s=1}^{S}$ is a set of **simulation draws**, each of length D, extracted from some **true** empirical **distribution**.

- In this setup, the contraction mapping works like in BLP.

$$\boldsymbol{\delta}_m^{h+1} = \boldsymbol{\delta}_m^{h} + \log\left(\mathbf{s}_m\right) - \log\left(\widehat{\boldsymbol{s}}\left(\boldsymbol{\delta}_m^{h}, \mathbf{p}_m, \mathbf{X}_m; \theta_2\right)\right)$$

The improved BLP framework (3/5)

- The GMM problem is once again linear in $\boldsymbol{\theta}_1$ and non-linear in $\boldsymbol{\theta}_2$. Using a Quasi-Newton optimizer with a user-supplied gradient, Nevo speeds up the outer loop search over $(\boldsymbol{\Pi}, \boldsymbol{\sigma})$.

- Nevo estimates his augmented BLP model on quarterly-level data (1988-1992) about 25 brands of ready-to-eat cereals in 65 cities. His definition of "market" is thus a city-quarter.

- Let $t = 1, \ldots, T$ index time. Nevo's data are rich enough to estimate product (and possibly city-level) **fixed effects** ξ_j.

$$\delta_{jmt} = \boldsymbol{x}_{jmt}^{\mathrm{T}} \boldsymbol{\beta} - \alpha P_{jmt} + \xi_j + \Delta\xi_{jmt}$$

- The distributions of $\boldsymbol{d}_{imt}$ used for simulating market shares are taken from a city's yearly Current Population Survey. In particular, Nevo uses the empirical distributions of income, age and number of children.

The improved BLP framework (4/5)

- Another important innovation by Nevo is that he includes a product j's **price in other cities** in the **instruments set**.

- This is necessary for him because the product characteristics $\boldsymbol{x}_{jmt}$ and cost shifters $\boldsymbol{w}_{jmt}$ have little statistical variation at the city-quarter level the data.

- The argument is that conditional on product fixed effects ξ_j, prices in other cities P_{jnt} are correlated to P_{jmt} for $n \neq m$, but are still exogenous in the sense that $\mathbb{E}\left[\Delta\xi_{jmt}\middle| P_{jnt}\right] = 0$.

- This allows to simplify the model by **removing the supply side**, and avoid assumptions about the cost function.

- This is important for the sake of Nevo's **research question**, which is about finding an explanation of the high **price-cost margins** (PMC) in this particular industry.

The improved BLP framework (5/5)

- With this shortcut, Nevo is able to **estimate** the **marginal cost** MC_{jmt} via the model estimates of demand elasticities and the following standard single-product pricing rule.

$$S_{jmt} = \left| \frac{\partial S_{jmt} \left(P_{jmt} \right)}{\partial P_{jmt}} \right| \left(P_{jmt} - MC_{jmt} \right)$$

- Nevo also considers extension of this pricing rule that allow for multi-product firms or collusive behavior *à la* Bresnahan (1987). Specifically, the general pricing rule at time t is:

$$\boldsymbol{s} \left(\boldsymbol{p}_t \right) = \left[\mathbf{H}_t \circ \boldsymbol{S} \left(\boldsymbol{p}_t \right) \right] \left(\boldsymbol{p}_t - \boldsymbol{c}_t \right)$$

  where $\boldsymbol{s} \left(\boldsymbol{p}_t \right)$ are the products' market shares as a function of prices $\boldsymbol{p}_t$, $\boldsymbol{c}_t$ is a vector of marginal costs, $\boldsymbol{S} \left(\boldsymbol{p}_t \right)$ is a matrix of price elasticities of size $J \times J$, and $\mathbf{H}_t$ is as in Bresnahan.

- The PMCs predicted by the single-product rule are closer to the observed PMCs: thus, Nevo rules out collusive behavior.

The frontier in demand estimation

While somewhat dated, the BLP model in its Nevo version is still the dominant framework for demand estimation. The extensions and applications that accumulated over time did not innovate the core methodology.

Present-day successful research that applies the BLP framework satisfies at least one of the following three conditions:

1. it addresses an important research question;
2. it complements BLP estimation with reduced form evidence;
3. it utilizes novel instruments in the GMM problem.

On the methodological side, interest for *non-parametric* methods that would let get away with the parametric BLP assumptions is currently growing; see e.g. the review by Berry and Haile (2021).

Estimation of nested demand structures (1/3)

A relevant extension of the workhorse BLP model is the **random coefficient nested logit** (RNCL) model originally proposed by Brenkers and Verboven (2006).

This model assumes that the ε_{ji} shocks have the same dependence structure leading to a nested logit model (Lecture 13), hence:

$$S_{\ell j}\left(\boldsymbol{\delta}, \boldsymbol{X}; \boldsymbol{\sigma}, \rho\right) = \int_{\mathbb{R}^K} \frac{\exp\left(\rho I_\ell\right)}{1 + \sum_{l=1}^{L} \exp\left(\rho I_l\right)} \cdot$$
$$\cdot \frac{\exp\left(\left(\delta_{\ell j} + \boldsymbol{x}_{\ell j}^{\mathrm{T}} \left(\mathbf{I}\boldsymbol{\sigma}\right) \boldsymbol{\upsilon}_s\right) / \rho\right)}{\sum_{k=1}^{J_\ell} \exp\left(\left(\delta_{\ell k} + \boldsymbol{x}_{\ell k}^{\mathrm{T}} \left(\mathbf{I}\boldsymbol{\sigma}\right) \boldsymbol{\upsilon}_s\right) / \rho\right)} f_{\boldsymbol{\nu}}\left(\boldsymbol{\nu}, \boldsymbol{X}; \boldsymbol{\sigma}\right) \mathrm{d}\boldsymbol{\nu}$$

where products are identified by **limbs** $\ell = 1, \ldots, L$ as well as **branches** $j = 1, \ldots, J_\ell$, and ρ is the "anti-correlation" parameter typical of the nested logit, taken here as constant across limbs.

Estimation of nested demand structures (2/3)

In the term $\exp\left(\rho I_\ell\right)\left[1+\sum_{l=1}^{L}\exp\left(\rho I_l\right)\right]^{-1}$, it is for $l = 1, \ldots, L$:

$$I_l = \log\left(\sum_{l=1}^{J_\ell}\exp\left(\left(\delta_{lk} + \boldsymbol{x}_{lk}^{\mathrm{T}}\left(\mathbf{I}\boldsymbol{\sigma}\right)\boldsymbol{v}_s\right)/\rho\right)\right)$$

which represents the "deterministic value" of limb l for a single consumer. Note the implicit inclusion of an outside option.

As already shown by Berry (1994), without random coefficients (that is, $\boldsymbol{\sigma} = \mathbf{0}$) this model becomes a standard nested logit and an analytic solution for $\delta_{\ell j}$ exists:

$$\begin{aligned}\delta_{\ell j} &= \log\left(S_{\ell j}\right) - \log\left(S_0\right) - \rho\log\left(S_{j|\ell}\right)\\ &= \boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta} - \alpha P_j + \xi_j\end{aligned}$$

where $S_{j|\ell}$ is the "conditional" share of j within limb ℓ.

Estimation of nested demand structures (3/3)

For reasons about interpretation of the model's parameters, the RNCL has become popular to study demand in markets with a clear nested structure of products, such as for example alcoholic beverages (e.g. "beer" versus "whiskey" as distinct limbs).

Estimation proceeds pretty much like in BLP, but with a major difference, as the contraction mapping must be "dampened" by parameter ρ (Grigolon and Verboven, 2014):

$$\boldsymbol{\delta}_m^{h+1} = \boldsymbol{\delta}_m^h + \rho \left[\log \left(\mathbf{s}_m \right) - \log \left(\widehat{\boldsymbol{s}} \left(\boldsymbol{\delta}_m^h, \mathbf{X}_m; \boldsymbol{\sigma} \right) \right) \right]$$

which is itself estimated in the outer loop. A problem can occur if ρ is close to zero, as the updates h become arbitrarily slow.

Intuitively, ρ close to zero implies high correlation (substitution) between products within a limb, and hence more "noise."

Current best practices in demand estimation

Originally, the BLP model was estimated via “quasi-raw” custom code (often in MATLAB). The “inner loop, outer loop” algorithm was very difficult to code from scratch (and it still is).

Nowadays, researchers rely on established libraries for languages such as Python or R; the PyBLP library for Python is especially popular. Practitioners identified over time some “best practices” to facilitate and improve BLP estimation.

These practices are reviewed by Conlon and Gortmaker (2020), the authors of PyBLP, and include:

1. projection methods to incorporate many fixed effects ξ_j;
2. techniques and “tricks” to facilitate numerical optimization;
3. evaluations of different instruments for efficiency’s sake.