Demand Estimation

Paolo Zacchia

Microeconometrics

Lecture 13

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 thus reviewed alongside other **multinomial** models.

- 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.

- Finally, the lecture overviews some 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:

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

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

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

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

which is an expression for the **budget share** $S_{ji}(\boldsymbol{p}, U_i)$ 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**:

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

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 $\ell = 1, \ldots, J$:

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

- …the **income elasticity**:

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

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

$$\eta^{P_\ell}_{Y^c_{ji}} \equiv \frac{P_\ell}{Y^c_{ji}(\boldsymbol{p}, U_i)} \frac{\partial Y^c_{ji}(\boldsymbol{p}, U_i)}{\partial P_\ell} = \frac{\partial \log Y^c_{ji}(\boldsymbol{p}, U_i)}{\partial \log P_\ell}$$

- …all related via the **Slutsky** equation $\eta^{P_\ell}_{Y_{ji}} = \eta^{P_\ell}_{Y^c_{ji}} - \eta^{M}_{Y_{ji}} M_i$.

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:

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

  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 $k = 1, \dots, J$.

- For $\ell = 1, \dots, J$, 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 $\ell = 1, \dots, J$, 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) that is summarized in Lecture 11, which however requires restrictive functional forms for demand.

- 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.

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.

Review of multinomial response models

What follows is an overview of leading econometric **multinomial** response models. The following are presented in sequence:

- the **multinomial logit model**;
- the **nested (multinomial) logit model**;
- the **mixed (multinomial) logit model**;
- the **multinomial probit model**;
- and **ordered multinomial models** (probit and logit).

Emphasis is on practical issues of estimation and implementation in practice, and interpretation of the resulting estimates.

The multinomial logit model (1/9)

- The **multinomial logit** is an important limited dependent variable (LDV) model for a **multinomial** outcome Y_i.

- That is, the support of Y_i (write it $\mathbb{Y}$) is *finite* and *countable*.

- Let there be J alternative realizations of Y_i ($|\mathbb{Y}| = J$).

- Typically, the dependent variable is coded over a collection of integers, $Y_i = 1, 2, \dots, J$: however, numbers do **not** imply an ordered relationship of any sort.

- Thus, the outcome variable can be conveniently re-coded in terms of J Bernoulli variables Y_{ji} for $j = 1, \dots, J$ with:

$$Y_{ji} = \begin{cases} 1 & \text{if } Y_i = j \\ 0 & \text{otherwise.} \end{cases}$$

The multinomial logit model (2/9)

- Interest in this model falls on the *probability* that any of the J possible realizations of Y_i occurs as a function of some K observable characteristics $\boldsymbol{x}_{ji} = (X_{1ji}, X_{2ji}, \dots, X_{Kji})$ that are possibly **specific to alternative** $j = 1, \dots, J$.

- If for example Y_i represents different product alternatives, $\boldsymbol{x}_{ji}$ may represent the subjective evaluation that a consumer makes of all these alternatives.

- Because this amounts to specifying *conditional* probabilities, the model is often called **conditional multinomial logit**.

- The (conditional) multinomial logit's defining feature is the following expression for the probability of all alternatives.

$$p_{ji} \equiv \mathbb{P}\left(Y_{ji} = 1 \middle| \boldsymbol{x}_{1i}, \dots, \boldsymbol{x}_{Ji}\right) = \frac{\exp\left(\boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta}\right)}{\sum_{k=1}^{J} \exp\left(\boldsymbol{x}_{ki}^{\mathrm{T}} \boldsymbol{\beta}\right)}$$

  where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_K)$ is a parameter vector of interest.

The multinomial logit model (3/9)

- Note that if $\boldsymbol{x}_{ji}$ were constant across the J alternatives, that is $\boldsymbol{x}_{1i} = \boldsymbol{x}_{2i} = \cdots = \boldsymbol{x}_{Ji} = \boldsymbol{x}_i$, this model would be moot: all the J choices would be equally likely.

- However, in this case one can re-formulate the model as:

$$p_{ji} \equiv \mathbb{P}\left(Y_{ji} = 1 \middle| \boldsymbol{x}_i\right) = \frac{\exp\left(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}_j\right)}{\sum_{k=1}^{J} \exp\left(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}_k\right)}$$

  where $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jK})$ is one out of J **alternative-specific** parameter vectors of interest.

- The inability to estimate J alternative-specific parameters if $\boldsymbol{x}_{ji}$ is not constant over j is an identification problem!

- Most typically, $\boldsymbol{x}_{ji}$ features both alternative-specific as well as "constant" characteristics. The elements of $\boldsymbol{\beta}_j$ associated with the former are constrained constant across alternatives.

The multinomial logit model (4/9)

These different levels of variation for the observed characteristics $\boldsymbol{x}_{ji}$ and for the parameters $\boldsymbol{\beta}_j$ led to a use of language that may appear confusing. Many researchers call:

- a plain **multinomial logit** a model that features fixed $\boldsymbol{x}_i$ and varying $\boldsymbol{\beta}_j$;
- an actual **conditional multinomial logit** a model that on the contrary features varying $\boldsymbol{x}_{ji}$ and fixed $\boldsymbol{\beta}$;
- a **mixed multinomial logit** a model that “mixes” both.

This specific use of terminology may appear rather confusing to econometricians, who are typically accustomed to call “mixed” a multinomial logit with *random parameters* (more on this later).

For simplicity, the following treatment sticks to the “conditional multinomial logit” with varying $\boldsymbol{x}_{ji}$ and fixed $\boldsymbol{\beta}$.

The multinomial logit model (5/9)

Make the following observations.

- One can always reformulate an alternative-invariant variable X_i as a vector of length J: $\boldsymbol{x}^*_{ji} = (D_{1ji}X_i, \ldots, D_{Jji}X_i)$; with $D_{\ell ji} = 1$ if $\ell = j$ and $D_{\ell ji} = 0$ otherwise, for $\ell = 1, \ldots, J$.

- Hence, the J parameters associated with $\boldsymbol{x}^*_{ji}$ correspond to alternative-specific parameters.

- If $\boldsymbol{x}_{ji}$ contains a "constant" vector that is thus dummified, its parameters are interpreted as the *realization probabilities* conditional on all other $\boldsymbol{x}_{ji}$'s being set at zero.

Although the "conditional multinomial logit" is more general, for the sake of practical implementation and estimates interpretation a researcher must always pay attention to the level of variation of the observable characteristics $\boldsymbol{x}_{ji}$'s.

The multinomial logit model (6/9)

Like all LDV models, the multinomial logit admits a structural interpretation in terms of **latent variables**. Let:

$$V_{ji} = \boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_{ji}$$

be the **utility** associated by observation i to the j-th alternative. Here ε_{ji} is a **random** component of the utility V_{ji}. It is assumed that alternative j is "chosen" by observation i if it is the one that delivers the highest utility.

$$Y_{ji} = 1 \;\Leftrightarrow\; V_{ji} = \max\left\{V_{1i}, \dots, V_{Ji}\right\}$$

Furthermore, if ε_{ji} is **i.i.d.** with

$$\varepsilon_{ji} \sim \text{Gumbel}\left(0, 1\right)$$

that is, the random component follows the Gumbel distribution with standard parameters, then the realization probabilities take the multinomial logit form, as it is shown next.

The multinomial logit model (7/9)

$$
\begin{aligned}
p_{ji} &= \mathbb{P}\left(\bigcup_{k\neq j}\{V_{ji} \geq V_{ki}\}\right) \\
&= \mathbb{P}\left(\bigcup_{k\neq j}\left\{\varepsilon_{ki} \leq \varepsilon_{ji} + \left(\boldsymbol{x}_{ji} - \boldsymbol{x}_{ki}\right)^{\mathrm{T}} \boldsymbol{\beta}\right\}\right) \\
&= \int_{-\infty}^{\infty} \prod_{k\neq j} \exp\left(-\exp\left(-\varepsilon_{ji} - \left(\boldsymbol{x}_{ji} - \boldsymbol{x}_{ki}\right)^{\mathrm{T}} \boldsymbol{\beta}\right)\right) \frac{\exp\left(-\varepsilon_{ji}\right)}{\exp\left(\exp\left(-\varepsilon_{ji}\right)\right)} d\varepsilon_{ji} \\
&= \int_{\infty}^{0} -\prod_{k\neq j} \exp\left(-u \exp\left(\left(\boldsymbol{x}_{ki} - \boldsymbol{x}_{ji}\right)^{\mathrm{T}} \boldsymbol{\beta}\right)\right) \frac{1}{\exp\left(u\right)} du \qquad \left[u = \exp\left(-\varepsilon_{ji}\right)\right] \\
&= \int_{0}^{\infty} \exp\left(-u\left[1 + \sum_{k\neq j} \exp\left(\left(\boldsymbol{x}_{ki} - \boldsymbol{x}_{ji}\right)^{\mathrm{T}} \boldsymbol{\beta}\right)\right]\right) du \\
&= \frac{1}{1 + \sum_{k\neq j} \exp\left(\left(\boldsymbol{x}_{ki} - \boldsymbol{x}_{ji}\right)^{\mathrm{T}} \boldsymbol{\beta}\right)} \\
&= \frac{\exp\left(\boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta}\right)}{\sum_{k=1}^{J} \exp\left(\boldsymbol{x}_{ki}^{\mathrm{T}} \boldsymbol{\beta}\right)}
\end{aligned}
$$

The multinomial logit model (8/9)

- At first the Gumbel assumption might seem rather arbitrary. Note though that for $j, k = 1, \ldots, J$:

$$(V_{ji} - V_{ki}) - (\boldsymbol{x}_{ji} - \boldsymbol{x}_{ki})^{\mathrm{T}} \boldsymbol{\beta} = \varepsilon_{ji} - \varepsilon_{ki} \sim \text{Logistic}\,(0, 1)$$

  the difference between any two random components follows the standard **logistic** distribution (Observation 14, Lecture 3) which can be thought as a more natural choice.

- If the scale parameter is **unrestricted**: $\varepsilon_{ji} \sim \text{Gumbel}\,(0, \sigma)$, the alternative-specific probabilities are hardly changed:

$$p_{ji} \equiv \mathbb{P}\left(Y_{ji} = 1 \middle| \boldsymbol{x}_{1i}, \ldots, \boldsymbol{x}_{Ji} \right) = \frac{\exp\left(\boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta} / \sigma\right)}{\sum_{k=1}^{J} \exp\left(\boldsymbol{x}_{ki}^{\mathrm{T}} \boldsymbol{\beta} / \sigma\right)}$$

  and consequently $\boldsymbol{\beta}$ and σ are not separately identified. This motivates the normalization $\sigma = 1$.

The multinomial logit model (9/9)

How to interpret the model's coefficients $\boldsymbol{\beta}$?

- They allow to calculate the **marginal effects** of changes in $\boldsymbol{x}_{ji}$ on the realization probability of each alternative.

$$\frac{\partial p_{ji}}{\partial \boldsymbol{x}_{ki}} = p_{ji} \left(\mathbb{1} \left[j = k \right] - p_{ki} \right) \boldsymbol{\beta}$$

  where p_{ki} is understood as a function of $(\boldsymbol{x}_{1i}, \dots, \boldsymbol{x}_{Ji})$ for all $k = 1, \dots, J$. Similarly to simpler logit and probit models, such marginal effects must be computed and/or averaged at specific realizations of $(\boldsymbol{x}_{1i}, \dots, \boldsymbol{x}_{Ji})$.

- Under the structural interpretation of the model, they also bear an interpretation in terms of **marginal utilities**.

$$\frac{\partial V_{ji}}{\partial \boldsymbol{x}_{ji}} = \frac{\partial \left(\boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_{ji} \right)}{\partial \boldsymbol{x}_{ji}} = \boldsymbol{\beta}$$

Estimation of the multinomial logit model (1/4)

- The likelihood function of this model is:

$$\mathcal{L}\left(\boldsymbol{\beta}\middle|\,\mathbf{x}_{1i},\ldots,\mathbf{x}_{Ji}\right)=\prod_{i=1}^{N}\prod_{j=1}^{J}p_{ji}^{y_{ji}}$$

  where p_{ji} is implicitly treated as a function of the *realizations* $(\mathbf{x}_{1i},\ldots,\mathbf{x}_{Ji})$ and y_{ji} is the realization of Y_{ji} for $j=1,\ldots,J$ (and it follows that $\sum_{j=1}^{J}y_{ji}=\sum_{j=1}^{J}Y_{ji}=1$).

- Thus, the log-likelihood function is as follows.

$$\log\mathcal{L}\left(\boldsymbol{\beta}\middle|\,\mathbf{x}_{1i},\ldots,\mathbf{x}_{Ji}\right)=\sum_{i=1}^{N}\sum_{j=1}^{J}y_{ji}\log\left(p_{ji}\right)$$

- Define the following quantity.

$$\bar{\mathbf{x}}_i=\sum_{j=1}^{J}p_{ji}\mathbf{x}_{ji}=\frac{\sum_{j=1}^{J}\exp\left(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta}\right)\mathbf{x}_{ji}}{\sum_{k=1}^{J}\exp\left(\mathbf{x}_{ki}^{\mathrm{T}}\boldsymbol{\beta}\right)}$$

Estimation of the multinomial logit model (2/4)

- The First Order Conditions are as follows:

$$\frac{\partial \log \mathcal{L}\left(\left.\boldsymbol{\beta}\right|\mathbf{x}_{1i},\dots,\mathbf{x}_{Ji}\right)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\sum_{j=1}^{J}\frac{y_{ji}}{p_{ji}}\frac{\partial p_{ji}}{\partial \boldsymbol{\beta}}$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ji}\left(\mathbf{x}_{ji}-\bar{\mathbf{x}}_i\right)$$
$$= \mathbf{0}$$

since $\partial p_{ji}/\partial \boldsymbol{\beta} = p_{ji}\left(\mathbf{x}_{ji}-\bar{\mathbf{x}}_i\right)$ as it is possible to verify. The parameters are "buried" within $\bar{\mathbf{x}}_i$. Since there is no closed form solution, the estimates are obtained numerically.

- Some further algebra yields the Hessian of the log-likelihood function, which may be useful for inference purposes.

$$\frac{\partial \log \mathcal{L}\left(\left.\boldsymbol{\beta}\right|\mathbf{x}_{1i},\dots,\mathbf{x}_{Ji}\right)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\mathrm{T}}} = -\sum_{i=1}^{N}\sum_{j=1}^{J} y_{ji}\left(\mathbf{x}_{ji}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{ji}-\bar{\mathbf{x}}_i\right)^{\mathrm{T}}$$

Estimation of the multinomial logit model (3/4)

- These equations differ for models that feature *only* constant characteristics $\mathbf{x}_i$ and varying parameters $\boldsymbol{\beta}_j$.

- In particular, the First Order Conditions for $\boldsymbol{\beta}_j$, $j = 1, \dots, J$ are as follows (yet again without closed form solution).

$$\frac{\partial \log \mathcal{L}\left(\boldsymbol{\beta} \middle| \mathbf{x}_i \right)}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{N} \left(y_{ji} - p_{ji} \right) \mathbf{x}_i = \mathbf{0}$$

- Recall that p_{ji} is a function of *all* the parameters! Note that for $k = 1, \dots, J$ it is $\partial p_{ji} / \partial \boldsymbol{\beta}_k = p_{ji} \left(\mathbb{1}\left[j = k \right] - p_{ki} \right) \mathbf{x}_i$.

- The Hessian of the log-likelihood function instead has blocks with the following form, for $j, k = 1, \dots, J$.

$$\frac{\partial \log \mathcal{L}\left(\boldsymbol{\beta} \middle| \mathbf{x}_i \right)}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_k^{\mathrm{T}}} = - \sum_{i=1}^{N} \sum_{j=1}^{J} p_{ji} \left(\mathbb{1}\left[j = k \right] - p_{ki} \right) \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}$$

Estimation of the multinomial logit model (4/4)

- Sometimes the alternatives **available** to single observations are not the same. Denote the choice set for observation i as $\mathcal{C}_i$. In such a case, the multinomial logit is still well defined. The likelihood function changes as:

$$\mathcal{L}\left(\boldsymbol{\beta}\middle|\,\mathbf{x}_{1i},\ldots,\mathbf{x}_{Ji}\right)=\prod_{i=1}^{N}\prod_{j\in\mathcal{C}_i}\left[\frac{\exp\left(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta}\right)}{\sum_{k\in\mathcal{C}_i}\exp\left(\mathbf{x}_{ki}^{\mathrm{T}}\boldsymbol{\beta}\right)}\right]^{y_{ji}}$$

  and estimation proceeds as in the standard case.

- In other cases, the choice set is so large as to make estimation impractical. McFadden (1978) showed that one can still get consistent estimates with a likelihood function like:

$$\mathcal{L}\left(\boldsymbol{\beta}\middle|\,\mathbf{x}_{1i},\ldots,\mathbf{x}_{Ji}\right)=\prod_{i=1}^{N}\prod_{j\in\mathcal{K}_i}\left[\frac{\exp\left(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta}\right)}{\sum_{k\in\mathcal{K}_i}\exp\left(\mathbf{x}_{ki}^{\mathrm{T}}\boldsymbol{\beta}\right)}\right]^{y_{ji}}$$

  where $\mathcal{K}_i$ is a **random subset** of alternatives associated to i that is selected so as to include i's realized outcome Y_i.

Independence of irrelevance alternatives

- The fundamental property of the multinomial logit model is the **independence of irrelevant alternatives** (IIA) that is featured by realization probabilities. In short:

$$\frac{p_{ji}}{p_{ki}} = \exp\left((\boldsymbol{x}_{ji} - \boldsymbol{x}_{ki})^{\mathrm{T}} \boldsymbol{\beta} \right)$$

  for any $j, k = 1, \dots, J$. Thus, for every observation pair the ratio between the realization probabilities of two alternatives is constant, and unaffected by other alternatives ℓ and their characteristics $\boldsymbol{x}_{\ell i}$.

- This may be **unrealistic** in many settings, as illustrated by the "red bus, blue bus" famous example (McFadden, 1974). Suppose that one is studying the determinants of choosing a "red bus" (j) against a car (k) as means of transportation. A two-outcomes model would return some ratio p_{ji}/p_{ki}. Then a "blue bus" (ℓ) is introduced. Realistically, p_{ki} should not vary, but IIA must be violated for $p_{ji} + p_{ki} + p_{\ell i} = 1$ to hold.

Limitations of the multinomial logit

The multinomial logit is extremely popular: it is based on simple expressions, it is easy enough to estimate, and it can be motivated in ways other than the Gumbel-distributed latent shocks ε_{ji}.

However, to an econometrician's eye it also features three major **limitations**.

1. The non-random "tastes" of individuals/observations, that is the $\boldsymbol{\beta}$ parameters, are unrealistically **homogeneous**.

2. As argued, the **substitution patterns** between alternatives are often unrealistic because of IIA.

3. The model is generally not well suited to data that feature autocorrelation in time or spatial correlation.

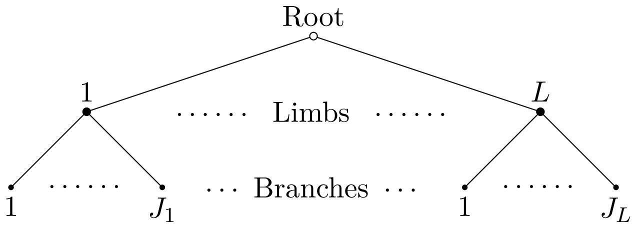
The next three multinomial models aim at addressing limitations 1. and 2. – the third one requires "Generalized Extreme Value" models, that are outside the scope of this review.

The nested logit model (1/3)

- It was McFadden himself (1978) who proposed an extension of the multinomial logit that addresses issues of IIA.

- In the **nested logit** the alternative outcomes have a "tree-like" **hierarchical structure**, with "limbs" and "branches." The J alternatives are thought as "branches" grouped across L "limbs;" each limb has J_ℓ branches with $\sum_{\ell=1}^{L} J_\ell = J$.

- Thus, alternatives are denoted by $Y_{\ell ji}$ with $j = 1, \dots, J_\ell$ and $\ell = 1, \dots, L$. They can be represented as follows.



The nested logit model (2/3)

- Let there be H **limb-specific** observable characteristics $\boldsymbol{z}_{\ell i}$, and K_l **branch-specific** $\boldsymbol{x}_{\ell j i}$ characteristics for $\ell = 1, \dots, L$.

- In the nested logit model, the realization probabilities are:

$$p_{\ell j i} = \underbrace{\frac{\exp\left(\boldsymbol{z}_{\ell i}^{\mathrm{T}} \boldsymbol{\alpha} + \rho_\ell I_\ell\right)}{\sum_{h=1}^{L} \exp\left(\boldsymbol{z}_{h i}^{\mathrm{T}} \boldsymbol{\alpha} + \rho_h I_h\right)}}_{= p_{\ell i} \equiv \mathbb{P}(Y_{\ell i} = 1 | \cdot)} \underbrace{\frac{\exp\left(\boldsymbol{x}_{\ell j i}^{\mathrm{T}} \boldsymbol{\beta}_\ell / \rho_\ell\right)}{\sum_{k=1}^{J_\ell} \exp\left(\boldsymbol{x}_{\ell k i}^{\mathrm{T}} \boldsymbol{\beta}_\ell / \rho_\ell\right)}}_{= p_{j i | \ell} \equiv \mathbb{P}\left(Y_{\ell j i} = 1 \middle| Y_{\ell i} = 1, \cdot\right)}$$

  where $Y_{\ell i} = 1$ denotes selection of the ℓ-th limb; whereas I_ℓ is defined as follows for $\ell = 1, \dots, L$.

$$I_\ell = \log\left(\sum_{k}^{J_\ell} \exp\left(\boldsymbol{x}_{\ell k i}^{\mathrm{T}} \boldsymbol{\beta}_\ell / \rho_\ell\right)\right)$$

- The model's parameters are $\theta = (\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_L, \rho_1, \dots, \rho_L)$. The nested structure operates through the $\boldsymbol{\rho} = (\rho_1, \dots, \rho_L)$ parameters: if $\boldsymbol{\rho} = \boldsymbol{\iota}$, this is a standard multinomial logit.

The nested logit model (3/3)

- The latent variable representation of the nested logit is:

$$V_{\ell ji} = \boldsymbol{z}_{\ell i}^{\mathrm{T}} \boldsymbol{\alpha} + \boldsymbol{x}_{\ell ji}^{\mathrm{T}} \boldsymbol{\beta}_{\ell} + \varepsilon_{\ell ji}$$

  where $\varepsilon_{\ell ji}$ follows a joint GEV distribution that features $\boldsymbol{\rho}$ as measures of within-limb anti-correlation (McFadden, 1978).

- The likelihood function is most succinctly expressed in terms of the various realization probabilities involved:

$$\mathcal{L}(\boldsymbol{\theta}|\cdot) = \prod_{i=1}^{N} \prod_{\ell=1}^{L} \left(p_{\ell i}^{y_{\ell i}} \prod_{j=1}^{J_{\ell}} p_{ji|\ell}^{y_{\ell ji}} \right)$$

  and so is the log-likelihood function to be *jointly* maximized.

$$\log \mathcal{L}(\boldsymbol{\theta}|\cdot) = \sum_{i=1}^{N} \sum_{\ell=1}^{L} \left[y_{\ell i} \log(p_{\ell i}) + \sum_{j=1}^{J_{\ell}} y_{\ell ji} \log\left(p_{ji|\ell}\right) \right]$$

- For convenience, one can **sequentially** estimate first I_{ℓ} and $\boldsymbol{\beta}_{\ell}/\rho_{\ell}$ in branches; and second, $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ in limbs.

The mixed logit model (1/2)

The **random parameters logit** model, also called **mixed logit** by econometricians, is based upon a representation of the latent random utility that features **heterogeneous** "tastes."

$$V_{ji} = \boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta}_i + \varepsilon_{ji}$$

The key feature is that the parameters $\boldsymbol{\beta}_i$ are observation-specific and treated as **random**, typically jointly normal.

$$\boldsymbol{\beta}_i \sim \mathcal{N}\left(\boldsymbol{\beta}, \boldsymbol{\Sigma}\right)$$

For $\boldsymbol{u}_i = \boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\beta}_i - \boldsymbol{\beta}\right)$, the model can be re-written as follows.

$$\begin{aligned} V_{ji} &= \boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta} + \upsilon_{ji} \\ \upsilon_{ji} &= \boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{u}_i + \varepsilon_{ji} \end{aligned}$$

The shock ε_{ji} is still assumed to be standard Gumbel distributed, and to be independent across observations and alternatives.

The mixed logit model (2/2)

- Notice that for $j \neq k$, $\mathbb{C}\text{ov}\left(v_{ji}, v_{ki} \middle| \boldsymbol{x}_{ji}, \boldsymbol{x}_{ki}\right) = \boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{x}_{ki}$: this introduces correlation between alternatives, defying IIA!

- The realization probabilities are as follows:

$$p_{ji} = \int_{\mathbb{R}^K} \frac{\exp\left(\boldsymbol{x}_{ji}^{\mathrm{T}}\left(\boldsymbol{\beta} + \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{u}_i\right)\right)}{\sum_{k=1}^{J} \exp\left(\boldsymbol{x}_{ki}^{\mathrm{T}}\left(\boldsymbol{\beta} + \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{u}_i\right)\right)} \phi\left(\mathbf{u}_i\right) \mathrm{d}\mathbf{u}_i$$

  where $\phi(\cdot)$ is the p.d.f. of the standard multivariate normal distribution. This integral has no closed form solution.

- This model is thus estimated by **MSL**, given a sample of S Monte Carlo draws $\{\mathbf{u}_s\}_{s=1}^{S}$; $\boldsymbol{\Sigma}$ is often restricted ex ante.

$$\left(\widehat{\boldsymbol{\beta}}_{MSL}, \widehat{\boldsymbol{\Sigma}}_{MSL}\right) =$$

$$= \underset{(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\arg\max} \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ji} \log \left[\frac{1}{S} \sum_{s=1}^{S} \frac{\exp\left(\boldsymbol{x}_{ji}^{\mathrm{T}}\left(\boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{\frac{1}{2}} \mathbf{u}_s\right)\right)}{\sum_{k=1}^{J} \exp\left(\boldsymbol{x}_{ki}^{\mathrm{T}}\left(\boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{\frac{1}{2}} \mathbf{u}_s\right)\right)} \right]$$

The multinomial probit model

The **multinomial probit** model is also based on the standard representation of the latent random utility:

$$V_{ji} = \boldsymbol{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_{ji}$$

but the random component $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \ldots, \varepsilon_{Ji})$ is jointly normally distributed: a more natural choice than GEV distributions.

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

Observe that if $\boldsymbol{\Sigma}$ is non-diagonal, the alternatives are correlated, like in the mixed logit. Moreover, IIA does not hold in this model.

For all its advantages, this model features quite a major problem: its realization probabilities can be very difficult to compute.

$$p_{ji} = \int_{\mathbb{R}^K} \prod_{k \neq j} \mathbb{1}\left(\mathbf{x}_{ji}^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_{ji} \geq \mathbf{x}_{ki}^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_{ki}\right) \frac{1}{|\boldsymbol{\Sigma}|} \phi\left(\boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\varepsilon}_i\right) \mathrm{d}\boldsymbol{\varepsilon}_i$$

Even simulation methods struggle to estimate this model quickly. Furthermore, identification requires careful restrictions on $\boldsymbol{\Sigma}$.

Ordered multinomial models (1/2)

- What if the alternatives are naturally **ordered** (for example, Y_i represents a ladder of a product's qualities)? The models reviewed thus far are unsuited to address the problem.

- The solution are the **ordered multinomial models** that posit a latent variable representation

$$Y_i^* = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i$$

that implies selection of the j-th alternative if it "passes" a certain associated **threshold** α_{j-1}, for $j = 1, \dots, J$.

$$Y_i = j \Leftrightarrow \alpha_{j-1} < Y_i^* \leq \alpha_j$$

- There are J thresholds $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)$ that are treated as **parameters** to be estimated, alongside $\boldsymbol{\beta}$.

- Note that the observable characteristics $\boldsymbol{x}_i$ only vary at the level of units of observation here.

Ordered multinomial models (2/2)

- Let $F_{\varepsilon|\boldsymbol{x}}\left(\varepsilon_i|\,\boldsymbol{x}_i\right)$ be the c.d.f. for ε_i given $\boldsymbol{x}_i$: for example, the standard normal $\Phi\left(\cdot\right)$ for the **ordered probit**, the standard logistic $\Lambda\left(\cdot\right)$ for the **ordered logit**, or others. Then:

$$\begin{aligned}
p_{ji} &\equiv \mathbb{P}\left(Y_i = j|\,\boldsymbol{x}_i\right)\\
&= \mathbb{P}\left(\alpha_{j-1} < Y_i^* \leq \alpha_j|\,\boldsymbol{x}_i\right)\\
&= \mathbb{P}\left(\alpha_{j-1} - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} < \varepsilon_i \leq \alpha_j - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\,\middle|\,\boldsymbol{x}_i\right)\\
&= F_{\varepsilon|\boldsymbol{x}}\left(\alpha_j - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\,\middle|\,\boldsymbol{x}_i\right) - F_{\varepsilon|\boldsymbol{x}}\left(\alpha_{j-1} - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\,\middle|\,\boldsymbol{x}_i\right)
\end{aligned}$$

- ... which enables MLE via a familiar log-likelihood function.

$$\log\mathcal{L}\left(\boldsymbol{\beta}|\,\mathbf{x}_i\right) = \sum_{i=1}^{N}\sum_{j=1}^{J} y_i \log\left(p_{ji}\right)$$

- The **marginal effects** obtain from the p.d.f.s $f_{\varepsilon|\boldsymbol{x}}\left(\varepsilon_i|\,\boldsymbol{x}_i\right)$.

$$\frac{p_{ji}}{\partial\boldsymbol{x}_i} = \left[f_{\varepsilon|\boldsymbol{x}}\left(\alpha_j - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\,\middle|\,\boldsymbol{x}_i\right) - f_{\varepsilon|\boldsymbol{x}}\left(\alpha_{j-1} - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\,\middle|\,\boldsymbol{x}_i\right)\right]\boldsymbol{\beta}$$

Instrumental variables for multinomial models

An alternative estimation approach for **unordered** multinomial models is based on **moment conditions** of the following form:

$$\mathbb{E}\left[Y_{ji} - p_{ji}\left(\boldsymbol{x}_{ji}; \boldsymbol{\theta}\right) \middle| \boldsymbol{z}_{ji}\right] = \mathbb{E}\left[\boldsymbol{z}_{ji}\left(Y_{ji} - p_{ji}\left(\boldsymbol{x}_{ji}; \boldsymbol{\theta}\right)\right)\right] = \mathbf{0}$$

for $j = 1, \ldots, J$. These moment conditions feature:

- $p_{ji}\left(\boldsymbol{x}_{ji}; \boldsymbol{\theta}\right)$: the realization probability for the j-th choice, as a function of the characteristics $\boldsymbol{x}_{ji}$ and some parameters $\boldsymbol{\theta}$; for example, this can be a multinomial probit *simulated* p_{ji};

- $\boldsymbol{z}_{ji}$: a vector of **instruments**; possibly it is $\boldsymbol{z}_{ji} = \boldsymbol{x}_{ji}$, more generally it includes a different/larger set of shifters.

If one suspects that the latent variable error ε_{ji} correlates with $\boldsymbol{x}_{ji}$ and $p_{ji}\left(\boldsymbol{x}_{ji}; \boldsymbol{\theta}\right)$ is correctly specified, estimating $\boldsymbol{\theta}$ using these moments in a (G)MM/MSM framework can be the sound choice. This is also helpful for computationally demanding models, like the probit. However, this is generally less efficient than MLE.

Multinomial models and demand estimation

Thanks to this long review of multinomial models, it is now easier to introduce the more modern approaches to demand estimation.

A brief intellectual history is thus sketched.

1. It starts with the equilibrium entry model by Bresnahan and Reiss (1991), estimated by **ordered probit**.

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**.

A simple entry model (1/6)

The model by Bresnahan and Reiss (1991) is not a proper exercise in "demand estimation," as it does not estimate a full-fledged set of cross-product price elasticities. Instead, it aims to uncover the structure of demand to make sense of **competition** in a market.

This simple **entry model** is notable as it already displays some key features of later contributions in demand estimation:

- estimation based on "aggregate" **market-level** data where econometricians observe the number N_i of *firms* selling some homogeneous good or service, and other local characteristics (but **no** prices, market shares, costs or margins);
- a latent variable specification that is derived from economic theory, and incorporates both **demand** and **supply**;
- the consequent specification of a multinomial model: in this particular case, an **ordered probit**.

A simple entry model (2/6)

The **demand** function in market i is given by:

$$Q_i = D\left(P_i; \boldsymbol{y}_i, \boldsymbol{z}_i\right) = d\left(P_i; \boldsymbol{z}_i\right) S\left(\boldsymbol{y}_i\right)$$

where Q_i and P_i are clearly quantity and price, and:

- $d\left(P_i; \boldsymbol{z}_i\right)$ is the demand of a **representative consumer**;
- $S\left(\boldsymbol{y}_i\right)$ is the **number of consumers**;
- $\boldsymbol{y}_i$ and $\boldsymbol{z}_i$ are **demographic variables** that affect demand.

On the **supply** side, firms have:

- **fixed costs** $F\left(\boldsymbol{w}_i\right) + B$;
- **marginal costs** $MC\left(q, \boldsymbol{w}_i\right)$;
- **average variable costs** $AVC\left(q, \boldsymbol{w}_i\right)$;

given a **per-firm quantity** q and local **cost shifters** $\boldsymbol{w}_i$.

A simple entry model (3/6)

The econometric model makes conclusions about competition by studying the relationship between the **number of firms** N_i and market **"size"** S_i.

To understand the underlying economic intuition, write a firm's **average profits** as:

$$\Pi_i = \left[P_{N_i} - AVC\left(q_{N_i}, \boldsymbol{w}_i\right)\right] d\left(P_{N_i}; \boldsymbol{z}_i\right) \frac{S\left(\boldsymbol{y}_i\right)}{N_i} - F\left(\boldsymbol{w}_i\right) - B_{N_i}$$

where some variables are indexed by N_i to highlight that these are affected by the structure of **competition** or successive entry. Then, define firms' **entry threshold** as:

$$s_{N_i} \equiv \frac{S_{N_i}\left(\boldsymbol{y}_i\right)}{N_i} = \frac{F\left(\boldsymbol{w}_i\right) + B_{N_i}}{\left[P_{N_i} - AVC\left(q_{N_i}, \boldsymbol{w}_i\right)\right] d\left(P_{N_i}; \boldsymbol{z}_i\right)}$$

that is, the market share that firms need to *at least break even* for given N_i ($\Pi_i = 0$). This function is **increasing** in N_i.

A simple entry model (4/6)

To enable econometric estimation, a **parameterized** expression for profits is necessary. For $\boldsymbol{x}_i = (\boldsymbol{w}_i, \boldsymbol{z}_i)$:

$$\Pi_i = S(\boldsymbol{y}_i; \boldsymbol{\lambda}) V_{N_i}(\boldsymbol{w}_i, \boldsymbol{z}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}) - F_{N_i}(\boldsymbol{w}_i; \boldsymbol{\gamma}, \boldsymbol{\delta}) + \varepsilon_i$$

where $\varepsilon_i \sim \mathcal{N}(0,1)$ and is independent across observations, and:

- the **market size** is specified as follows:

$$S(\boldsymbol{y}_i; \boldsymbol{\lambda}) = \boldsymbol{y}_i^{\mathrm{T}} \boldsymbol{\lambda}$$

- firms' **variable profits** (for $S = 1$) are specified as follows:

$$V_{N_i}(\boldsymbol{x}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_1 - \sum_{n=2}^{N_i} \alpha_n + \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}$$

- and firms' **fixed costs** are specified as follows.

$$F_{N_i}(\boldsymbol{w}_i; \boldsymbol{\gamma}, \boldsymbol{\delta}) = \gamma_1 + \sum_{n=2}^{N_i} \gamma_n + \boldsymbol{w}_i^{\mathrm{T}} \boldsymbol{\delta}$$

A simple entry model (5/6)

- This yields in an estimable **ordered probit** where N_i is the outcome variable and Π_i is the latent variable.

- The ordered probit **thresholds** are the $\boldsymbol{\gamma}$ parameters; there are also "extra" thresholds $\boldsymbol{\alpha}$ that **interact** with $\boldsymbol{y}_i$. These parameters are meant to capture the effect of **competition**.

- To ensure identification, Bresnahan and Reiss normalize one element of $\boldsymbol{\lambda}$ (the parameter for total population) to one.

- The **entry thresholds** can be obtained from the estimates.

$$\widehat{S}_N = \frac{\widehat{\gamma}_1 - \sum_{n=2}^{N_i} \widehat{\gamma}_n + \bar{\mathbf{w}}^{\mathrm{T}} \widehat{\boldsymbol{\delta}}}{\widehat{\alpha}_1 - \sum_{n=2}^{N_i} \widehat{\alpha}_n + \bar{\mathbf{x}}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}}$$

- These are obtained from sample averages $\bar{\mathbf{w}}$ and $\bar{\mathbf{x}}$, and allow to calculate $\widehat{s}_N = \widehat{S}_N / N$ for any integer N.

A simple entry model (6/6)

- Bresnahan and Reiss estimate their model in several markets for local services: doctors, dentists, druggists (pharmacists), plumbers, tire dealers.

- They focus on several **isolated towns**. Note: heterogeneity between different doctors, dentists etc. is incorporated in ε_i.

- The results are meaningful: the econometric "thresholds" α_n and γ_n are hardly significant for $n = 3, 4, 5, \ldots$

- Suggestively, these markets approach **perfect competition** quite quickly as the number of oligopolists becomes modest.

- This is confirmed by the analysis and statistical tests about the ratios $\widehat{s}_N/\widehat{s}_M$ for $M < N$ (especially $M = 1$).

- This is a striking confirmation of **oligopoly theory**!

Towards the workhorse BLP framework

- Although it was not a proper exercise in demand estimation, the model by Bresnahan and Reiss showed the usefulness of multinomial models for market data.

- This stimulated more work on extensions of the multinomial logit (baseline, nested, mixed) for the case of interest.

- 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.

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 $k = 1, \ldots, K$, 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 $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$; 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:

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

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 $k = 0, 1, \ldots, J$:

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

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

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

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

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

the solution for mean utilities is straightforward in this case.

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

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:

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

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 (1/8)

- 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 (1982) 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 (2/8)

- 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** $T(\boldsymbol{\delta}; \boldsymbol{\sigma}) : \mathbb{R}^J \to \mathbb{R}^J$ 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 (3/8)

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

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

- Write the following **moment conditions**:

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

  also written more compactly as $\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{q}_{jm};\boldsymbol{\theta}\right)\right] = \boldsymbol{0}$, 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 $2L \geq |\boldsymbol{\theta}|$.

The workhorse BLP framework (4/8)

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

$$\widehat{\boldsymbol{\theta}}_{BLP} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[\overline{\mathbf{g}}_{M,J}\left(\mathbf{q}_{jm};\boldsymbol{\theta}\right)\right]^{\mathrm{T}} \mathbf{A}_{2L} \left[\overline{\mathbf{g}}_{M,J}\left(\mathbf{q}_{jm};\boldsymbol{\theta}\right)\right]$$

with $\overline{\mathbf{g}}_{M,J}\left(\mathbf{q}_{jm};\boldsymbol{\theta}\right) \equiv \frac{1}{MJ}\sum_{m=1}^{M}\sum_{j=1}^{J}\boldsymbol{g}\left(p_{jm},\mathbf{s}_m,\mathbf{w}_{jm},\mathbf{x}_{jm},\mathbf{z}_{jm};\boldsymbol{\theta}\right).$

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

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

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

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

2. in the **outer loop**, given $\widehat{\boldsymbol{\delta}}_m$, the particular value of $\boldsymbol{\theta}$ that minimizes the GMM objective function is sought.

The workhorse BLP framework (5/8)

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.

- 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 $\mathbf{A}_{2L}$ 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 (6/8)

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:

$$\boldsymbol{z}_{(kjm)} = \begin{pmatrix} X_{kjm} & \sum_{\ell \neq j, \ell \in \mathcal{J}_{jm}} X_{k\ell m} & \sum_{\ell \neq \mathcal{J}_{jm}} X_{k\ell m} \end{pmatrix}$$

  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 (7/8)

- 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 (8/8)

- 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 $\boldsymbol{\upsilon}_i = (\upsilon_{Pi}, \upsilon_{1i}, \ldots, \upsilon_{Ki})$ like in BLP (with the addition of υ_{Pi}), $\boldsymbol{d}_i = (d_{1i}, \ldots, d_{Di})$ 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 $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ is as before while $\boldsymbol{\theta}_2 = (\boldsymbol{\Pi}, \boldsymbol{\Sigma})$. Usually $\boldsymbol{\Sigma}$ is restricted to the diagonal like in BLP: $\boldsymbol{\Sigma} = \mathbf{I}\boldsymbol{\sigma}$.

- 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:

$$\widehat{S}_j\left(\boldsymbol{\delta}, \boldsymbol{p}, \boldsymbol{X}; \boldsymbol{\theta}_2\right) = \frac{1}{S}\sum_{s=1}^{S} \frac{\exp\left(\delta_j + \boldsymbol{r}_j^{\mathrm{T}}\left[\boldsymbol{\Pi}\mathbf{d}_s + \boldsymbol{\Sigma}\boldsymbol{\upsilon}_s\right]\right)}{\sum_{k=1}^{J}\exp\left(\delta_k + \boldsymbol{r}_k^{\mathrm{T}}\left[\boldsymbol{\Pi}\mathbf{d}_s + \boldsymbol{\Sigma}\boldsymbol{\upsilon}_s\right]\right)}$$

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.

$$\boldsymbol{\delta}_m^{h+1} = \boldsymbol{\delta}_m^{h} + \log\left(\mathbf{s}_m\right) - \log\left(\widehat{\boldsymbol{s}}\left(\boldsymbol{\delta}_m^{h}, \mathbf{p}_m, \mathbf{X}_m; \boldsymbol{\theta}_2\right)\right)$$

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.

See e.g. the paper by Miller and Weinberg (2017) about the effect of joint ventures for a successful example in all three domains.

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).