

# 1 Models of Demand for Differentiated Products

## 1.1 Basics

Differentiated goods set  $\mathcal{J}$  with  $\text{card}(\mathcal{J}) = J$ ; price vector  $\mathbf{p} = \{p_1, \dots, p_J\}^\top$  and market share vector  $\mathbf{s} = \{s_1, \dots, s_J\}^\top$ ;  $\mathbf{x}_j$  is the vector of good  $j$  characteristics. Good  $j = 0$  is the *outside good*, a construct used to allow substitution away from all goods in  $\mathcal{J}$ .

$\mathbf{q} \equiv \mathbf{q}(\mathbf{p}, \theta)$  is the demand system parameterised by  $\theta$ ; the functional form of  $\mathbf{q}$  is chosen such that it is **flexible** (Lau, 1986: “price elasticities are capable of arbitrary values subject only to the constraint of theoretical consistency”), yet estimable: for example, it is not possible to estimate an unrestricted  $J \times J$  matrix of cross-price effects with  $J$  observations, hence the need for **restrictions on the functional form** (which limit flexibility).

## 1.2 Parametric Assumptions and the Mean Utility Index

$$u_{ij} = (\beta + \nu_i)\mathbf{x}_j - \alpha p_j + \xi_j + \varepsilon_{ij} \quad (1)$$

- $\beta$  is the marginal utility from characteristics  $\mathbf{x}_j \in \mathbb{R}^K$ .
- $\nu_i$  is consumer  $i$ 's taste deviation, from a pre-specified distribution  $H(\mu_\nu, \sigma_\nu)$
- $\alpha$  is the marginal utility of money
- $\xi_j$  is the mean utility from  $j$ 's **unobserved** characteristics
- $\varepsilon_{it}$  is a random valuation of  $j$  across  $i$  with distribution  $G(\mu_\varepsilon, \sigma_\varepsilon)$

Equation (1) can be decomposed in individual- and good-specific components:

$$u_{ij} = \underbrace{\beta\mathbf{x}_j - \alpha p_j + \xi_j}_{\delta_j} + \nu_i\mathbf{x}_j + \varepsilon_{ij} \quad (2)$$

Since we only care about the ordinal, rather than the cardinal, implications of our utility function (consumers choose based on the *ranking* of goods, not on their utility ‘number’), the utility of all goods can be normalised based on the utility of a *numéraire* good. In this case, it is particularly convenient to take the **excluded good**  $j = 0$  as **numéraire**: in particular, we make the relatively harmless assumption that  $u_{i0} = 0 + \varepsilon_{i0}$ .

In general, market shares will be defined as

$$\mathbf{s}(\boldsymbol{\delta}, \boldsymbol{\sigma}) = \{s_1(\boldsymbol{\delta}, \boldsymbol{\sigma}), \dots, s_J(\boldsymbol{\delta}, \boldsymbol{\sigma})\} \text{ where } s_j(\boldsymbol{\delta}, \boldsymbol{\sigma}) = \Pr(u_{ij} > u_{ik} \ \forall k \neq j) \quad (3)$$

and, assuming for the moment that  $\nu_i = 0 \ \forall i$ ,

$$\begin{aligned} \Pr(u_{ij} > u_{ik} \ \forall k \neq j) &= \int_{\varepsilon} \Pr(\varepsilon_{ik} < \delta_j - \delta_k + \varepsilon_{ij} \ \forall k \neq j | \varepsilon_{ij}) g(\varepsilon_{ij}) d\varepsilon_{ij} \\ &\stackrel{i.i.d.}{=} \int_{\varepsilon} \prod_{j \neq k} G(\delta_j - \delta_k + \varepsilon_{ij}) g(\varepsilon) d\varepsilon \end{aligned} \quad (4)$$

which shows the necessity for assumptions on  $G(\cdot)$  to identify parameters in  $\delta(\cdot)$ .

---

These notes are based on Howard Smith's lectures from the 2018-2019 academic year. Of course, this is our interpretation of the material Howard presented. Good content is his; mistakes are ours.

### 1.3 Probit

$\{\varepsilon_{i1}, \dots, \varepsilon_{iJ}\} \sim N(0, \Sigma)$ . The probability of choosing  $j$  is

$$\Pr(u_{ij} > u_{ik} \forall k \neq j) = \int_{\varepsilon} \mathbb{1}\{\delta_j + \varepsilon_{ij} > \delta_k + \varepsilon_{ik}, \forall k \neq j\} f(\varepsilon) d\varepsilon \quad (5)$$

where  $f(\cdot)$  is a multivariate Normal density. The issue with Probit is that the integral must be simulated and the number of parameters in  $\Sigma$  might be too many if the matrix is left unrestricted ( $\frac{J(J-1)}{2}$  in principle).

### 1.4 Logit

The classic and most widely used assumption is that  $\varepsilon_{ij}$  is i.i.d. Extreme Value Type 1 (EVT-1) distributed. The EVT-1 distribution is characterised as

$$G^{\text{EVT1}}(x) = \exp(-\exp(-x)) \quad g^{\text{EVT1}}(x) = \exp(-x) \exp(-\exp(-x))$$

for which  $\sigma_{\varepsilon}$  does not have to be estimated, as mean and variance are fixed.

Equation (4) under EVT-1 is

$$s_j(\delta, \sigma) = \int_{\varepsilon} \prod_{j \neq k} e^{-e^{-(\delta_j - \delta_k + \varepsilon_{ij})}} e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon}} d\varepsilon = \int_{\varepsilon} \prod_j e^{-e^{-(\delta_j - \delta_k + \varepsilon_{ij})}} e^{-\varepsilon_{ij}} d\varepsilon$$

which, after some tedious derivation involving a change of variable, implies<sup>1</sup>

$$s_j(\delta, \sigma) = \frac{e^{\delta_j}}{\sum_{k \in \mathcal{J}} e^{\delta_k}} \quad (6)$$

Importantly, the distributional assumption of the logit model implies the following for **expected maximum utility** from choice set  $\mathcal{J}$ :

$$\mathbb{E}_{\varepsilon}[\max_{j \in \mathcal{J}}(\delta_j + \varepsilon_{ij})] = \ln \sum_{j \in \mathcal{J}} e^{\delta_j} \quad (7)$$

While computationally attractive, the logit model carries a number of theoretically unattractive implications:

1. **All new products increase expected utility.** From (7),  $e^{\delta_j} > 0$  for any value of  $\delta_j$ .
2. **Independence of Irrelevant Alternatives.**  $\frac{s_k}{s_j}$ ,  $\forall k, j \in \mathcal{J}$  depends only on the characteristics of the two goods, and hence the relative popularity of two goods is **by assumption** unaffected by changes in  $\mathcal{J}$  and by the characteristics of all goods that are not  $k, j$ .
3. **Inflexible price elasticities:**  $\frac{\partial s_j}{\partial p_k} = \frac{\partial s_j}{\partial \delta_k} \frac{\partial \delta_k}{\partial p_k} = -\frac{e^{\delta_j + \delta_k}}{(\sum_{\ell \in \mathcal{J}} e^{\delta_{\ell}})^2} (-\alpha) = \alpha s_j s_k$  and  $\frac{\partial s_j}{\partial p_k} \frac{p_k}{s_j} = \alpha p_k s_k$ .

### 1.5 Nested Logit

Suppose  $\mathcal{J}$  can be partitioned in  $r = 1, \dots, R$  groups  $\mathcal{J}_r$ , with each good in only one group. We assume that the outside good  $j = 0$  sits alone in its outside good group  $r = 0$ .

---

<sup>1</sup>The same holds for an arbitrary subset  $\mathcal{J}'$  of  $\mathcal{J}$ .

The error terms are conveniently assumed to have the following Generalised Extreme Value distribution: for  $j \in \mathcal{J}_r$ , errors for each group  $r$  are drawn from

$$G(\varepsilon_{i1}, \dots, \varepsilon_{iJ_r}) = \exp \left[ - \sum_r^R \left( \sum_{j \in \mathcal{J}_r} e^{-\frac{\varepsilon_{ij}}{1-\sigma_r}} \right)^{1-\sigma_r} \right] \quad (8)$$

where  $\sigma_r \in (0, 1)$ . If  $\sigma_r = 0 \forall r$ , the model boils down to the Logit model presented above.

As noted in Berry (1994), the GEV disturbance can be decomposed as the sum of two terms, a group-specific error and a within-group Logit error:

$$u_{ij}^{GEV} = \delta_j + \varepsilon_{ir} + (1 - \sigma_r)\varepsilon_{ij|r} \quad (9)$$

which implies that market share of good  $j$  **within group**  $r$  is defined as in the Logit model:

$$s_{j \in \mathcal{J}_r} = \Pr(u_{ij}^{GEV} > u_{ik}^{GEV} \forall k \in \mathcal{J}_r) \stackrel{i.i.d.}{=} \int_{\varepsilon} \prod_{k \neq j \in \mathcal{J}_r} G^{\text{EVT1}}\left(\frac{\delta_j - \delta_k}{1 - \sigma_r} + \varepsilon_{ij}\right) g^{\text{EVT1}}(\varepsilon) d\varepsilon \quad (10)$$

and thus the within-group Logit behaviour implies

$$s_{j \in \mathcal{J}_r} = \frac{s_j}{s_r} = \frac{e^{\frac{\delta_j}{1-\sigma_r}}}{\sum_{\ell \in \mathcal{J}_r} e^{\frac{\delta_\ell}{1-\sigma_r}}} \quad (11)$$

Furthermore, it can be shown that the model implies a choice **between groups** that also follows a Logit form:

$$s_r = \frac{\left( \sum_{\ell \in \mathcal{J}_r} e^{\frac{\delta_\ell}{1-\sigma_r}} \right)^{1-\sigma_r}}{\sum_{r' \in R} \left( \sum_{\ell' \in \mathcal{J}_{r'}} e^{\frac{\delta_{\ell'}}{1-\sigma_{r'}}} \right)^{1-\sigma_{r'}}} \quad (12)$$

whence the name **Nested Logit**: the model is effectively a two-tier Logit model. The share of good  $j$  in the **total market** is defined from (11)

$$s_j \equiv s_{j \in \mathcal{J}_r} s_r = \frac{e^{\frac{\delta_j}{1-\sigma_r}}}{\sum_{\ell \in \mathcal{J}_r} e^{\frac{\delta_\ell}{1-\sigma_r}}} s_r \quad (13)$$

Nested Logit relaxes the assumptions of Logit in a very *ad hoc* way: for example, **IIA is relaxed for goods that do not belong to the same group**; it still holds for good belonging to the same group. However, **groups not directly compared do not matter** - IIA at the group level.

## 1.6 (Random Coefficients) Mixed Logit

$$u_{ij} = (\beta + \sigma_\nu \nu_i) \mathbf{x}_j - \alpha p_j + \xi_j + \varepsilon_{ij} \quad (14)$$

where  $\varepsilon_{ij}$  is EVT1-distributed and  $\nu_i \sim H(\cdot)$  (usually normally distributed) with scaling parameter  $\sigma_\nu$  to be estimated.

Conditional on  $\nu_i$ , we have probabilities defined as in Logit, and market shares obtained by integrating over  $\nu_i$ :

$$\Pr(u_{ij} > u_{ik} \forall k \neq j | \nu_i) = \frac{e^{\delta_j + \sigma_\nu \nu_i \mathbf{x}_j}}{\sum_{\ell \in \mathcal{J}} e^{\delta_\ell + \sigma_\nu \nu_i \mathbf{x}_\ell}} \Rightarrow s_j(\boldsymbol{\delta}, \boldsymbol{\sigma}) = \int_{\nu} \frac{e^{\delta_j + \sigma_\nu \nu_i \mathbf{x}_j}}{\sum_{\ell \in \mathcal{J}} e^{\delta_\ell + \sigma_\nu \nu_i \mathbf{x}_\ell}} h(\nu) d\nu \quad (15)$$

which, since no analytical solution is available for usual specifications of  $H(\cdot)$ , requires simulation as

$$\tilde{s}_j(\boldsymbol{\delta}, \boldsymbol{\sigma}) = \frac{1}{T} \sum_t \frac{e^{\delta_j + \sigma_j \nu_j \mathbf{x}_j}}{\sum_{\ell \in \mathcal{J}} e^{\delta_\ell + \sigma_\ell \nu_\ell \mathbf{x}_\ell}} \quad (16)$$

where  $t$  denotes draw  $t$  out of  $T$  from the theoretical support of  $h(\nu)$ .

## 2 Estimation

### 2.1 Inversion of the Market Share Function

Discrete choice models can be generally written as

$$s_j = s_j(\mathbf{x}, \mathbf{p}, \boldsymbol{\xi}; \theta) = s_j(\boldsymbol{\delta}, \sigma)$$

where  $\theta$  is a vector of parameters and the last equality follows from the Mean Utility Index formulation. Note  $s_j$  is increasing in  $\xi_j$  and decreasing in  $\xi_k \forall k \neq j$ .

For estimation, given  $s_j(\boldsymbol{\delta}, \sigma)$ , we aim to separate  $\xi_j$  as a **residual** to be used in defining the identifying conditions.

This requires **invertibility** of  $s(\cdot)$ : if  $s(\boldsymbol{\delta}, \sigma)$  is invertible, there exists a unique  $\delta(\mathbf{s}, \sigma)$  and the Mean Utility Index formulation implies  $\delta(\mathbf{s}, \sigma) - (\beta \mathbf{x} - \alpha p_j) = \boldsymbol{\xi}(\mathbf{s}, \sigma)$ .

Berry (1994) establishes the following **conditions for invertibility of the share function**:<sup>2</sup>

1.  $\mathbf{s}(\boldsymbol{\delta}, \sigma)$  is differentiable in  $\boldsymbol{\delta}$ . This is generally satisfied by models with a continuous random utility  $(u_{ij} - \delta_j)$ , as all the ones we consider;
2.  $\mathbf{s}(\boldsymbol{\delta}, \sigma)$  obeys  $\frac{\partial s_j}{\partial \delta_j} > 0$  and  $\frac{\partial s_j}{\partial \delta_k} < 0 \forall j \neq k$ , which is implicit in the modelling of substitute choices;
3.  $s_j(\boldsymbol{\delta}, \sigma) \rightarrow 1$  as  $\delta_j \rightarrow \infty$  and  $s_j(\boldsymbol{\delta}, \sigma) \rightarrow 0$  as  $\delta_j \rightarrow -\infty$ .

under these conditions, there is a unique inversion  $\delta(\mathbf{s}, \sigma)$  yielding residual  $\boldsymbol{\xi}(\mathbf{s}, \sigma) = \delta(\mathbf{s}, \sigma) - (\beta \mathbf{x}_j - \alpha p_j)$ . Let us show that these conditions apply in the models presented above.

#### 2.1.1 Inversion of Logit

$s_j$  is 1. differentiable in  $\delta_j$ ; 2. increasing in  $\delta_j$  and decreasing in  $\delta_k$ ; 3. obedient to  $\delta$ -limit conditions.

$$s_j(\boldsymbol{\delta}, \sigma) = \frac{e^{\delta_j}}{\sum_{k \in \mathcal{J}} e^{\delta_k}} \Leftrightarrow \ln s_j(\boldsymbol{\delta}, \sigma) = \delta_j - \ln \sum_{k \in \mathcal{J}} e^{\delta_k}$$

hence (note that in this case there are no non-linear parameters  $\boldsymbol{\sigma}$ )

$$\delta_j = \beta \mathbf{x}_j - \alpha p_j + \xi_j = \ln s_j(\boldsymbol{\delta}) - \ln s_0(\boldsymbol{\delta}) \Leftrightarrow \xi_j = \ln s_j(\boldsymbol{\delta}) - \ln s_0(\boldsymbol{\delta}) - \beta \mathbf{x}_j + \alpha p_j \quad (17)$$

---

<sup>2</sup>Berry notes that conditions 1 and 2 are *sufficient* for inversion; condition 3 seems to follow from condition 2.

### 2.1.2 Inversion of Nested Logit

Just as in the case of Logit, the requirements are satisfied. Then from (13)

$$\ln s_j = \ln s_{j \in \mathcal{J}_r} + \ln s_r = \frac{\delta_j}{1 - \sigma_r} - \sigma_r \ln \left( \sum_{\ell \in \mathcal{J}_r} e^{\frac{\delta_\ell}{1 - \sigma_r}} \right) - \ln \sum_{r' \in R} \left( \sum_{\ell' \in \mathcal{J}_{r'}} e^{\frac{\delta_{\ell'}}{1 - \sigma_{r'}}} \right)^{1 - \sigma_{r'}}$$

and

$$\ln s_0 = \underbrace{\frac{\delta_0}{1 - \sigma_0}}_{=0 \text{ as } \delta_0=0} - \ln \sum_{r' \in R} \left( \sum_{\ell' \in \mathcal{J}_{r'}} e^{\frac{\delta_{\ell'}}{1 - \sigma_{r'}}} \right)^{1 - \sigma_{r'}}$$

one obtains

$$\ln s_j - \ln s_0 = \frac{\delta_j}{1 - \sigma_r} - \sigma_r \ln \left( \sum_{\ell \in \mathcal{J}_r} e^{\frac{\delta_\ell}{1 - \sigma_r}} \right)$$

and, noting that  $s_r = s_0 \times \left( \sum_{\ell \in \mathcal{J}_r} e^{\frac{\delta_\ell}{1 - \sigma_r}} \right)^{1 - \sigma_r}$ , we have

$$\begin{aligned} \ln s_j - \ln s_0 &= \frac{\delta_j}{1 - \sigma_r} - \frac{\sigma_r}{1 - \sigma_r} (\ln s_r - \ln s_0) \\ (1 - \sigma_r)(\ln s_j - \ln s_0) &= \delta_j - \sigma_r (\ln s_r - \ln s_0) \Leftrightarrow \ln s_j - \ln s_0 = \delta_j - \sigma_r (\ln s_r - \ln s_j) \end{aligned} \quad (18)$$

and

$$\xi_j = \delta_j - \beta \mathbf{x}_j + \alpha p_j = \ln s_j - \ln s_0 - \sigma_r (\ln s_r - \ln s_j) - \beta \mathbf{x}_j + \alpha p_j \quad (19)$$

### 2.1.3 Inversion of Probit and RCML

While no analytical form is available for  $\delta(\mathbf{s}, \boldsymbol{\sigma})$ , Berry Levinsohn and Pakes (1995) show that the following is a contraction mapping:

$$\boldsymbol{\delta}_t = \boldsymbol{\delta}_{t-1} + \ln \mathbf{s} - \ln \tilde{\mathbf{s}}(\boldsymbol{\delta}_{t-1}, \boldsymbol{\sigma}) \quad (20)$$

where  $t$  indexes loop iterations and  $\tilde{\mathbf{s}}(\cdot)$  is the computed share from previous loop (or from the initial guess if  $t = 1$ ).

Being a contraction mapping, the sequence  $\{\delta, d(\delta), d(d(\delta)), \dots\}$  converges to a unique fixed point  $\delta = d(\delta)$ . The contraction mapping returns a vector  $\boldsymbol{\delta}(\mathbf{s}, \boldsymbol{\sigma})$  for which  $\boldsymbol{\xi}(\mathbf{s}, \theta) = \boldsymbol{\delta}(\mathbf{s}, \boldsymbol{\sigma}) - (\beta \mathbf{x} - \alpha \mathbf{p})$ . Note that now  $\boldsymbol{\delta}(\cdot)$  depends on all market shares.

## 2.2 Estimation with Market-Level Data: Demand Side

Data available on market shares  $s_j$ , product characteristics  $\mathbf{x}_j$ , and prices  $p_j$  for  $j \in \mathcal{J}$ . Can have data for multiple markets (periods and/or cities) but not strictly necessary.  $s_0$  being a construct, one usually needs to impose some assumption on it.

Estimation proceeds by GMM: given residuals

$$\boldsymbol{\xi}(\theta) = \boldsymbol{\delta}(\mathbf{s}, \boldsymbol{\sigma}) - \beta \mathbf{x} + \alpha \mathbf{p} \quad (21)$$

one defines identifying restrictions as a function of instruments  $\mathbf{Z}$  indexed by  $m$  assumed orthogonal

to  $\xi$ , i.e.

$$\mathbb{E}[Z_m \xi(\theta_0)] = 0 \text{ with sample counterpart } \vartheta_m(\theta) \equiv \frac{1}{J} \sum_j^J [Z_{mj} \xi_j(\theta)] = 0 \quad (22)$$

stacking these moments as  $\vartheta(\theta) = \{\vartheta_1(\theta), \dots, \vartheta_M(\theta)\}^\top$  the GMM estimator is

$$\hat{\theta} \in \arg \min_{\theta \in \Theta} \vartheta(\theta)^\top W_J \vartheta(\theta) \quad (23)$$

with a  $W_J$  weighting matrix.

### 2.2.1 GMM for Logit and Nested Logit

For these two models we have

$$\begin{aligned} \xi_j &= \ln s_j - \ln s_0 - \beta \mathbf{x}_j + \alpha p_j \\ \xi_j &= \ln s_j - \ln s_0 - \beta \mathbf{x}_j + \alpha p_j + \sigma_r \underbrace{(\ln s_j - \ln s_r)}_{\equiv \ln s_{j \in \mathcal{J}_r}} \end{aligned}$$

hence, defining  $Y_j \equiv \ln s_j - \ln s_0$ ,  $X_j = [\mathbf{x}_j, p_j]$  for Logit, and  $X_j = [\mathbf{x}_j, p_j, \ln s_{j \in \mathcal{J}_r}]$  for Nested Logit, one has the linear GMM estimator

$$\hat{\theta} = [X^\top Z W_J Z^\top X]^{-1} [X^\top Z W_J Z^\top Y] \quad (24)$$

### 2.2.2 GMM for Probit and RMCL

As an analytical solution is not available, we solve by iteration:

1. Start from an initial vector  $\theta_0 = (\sigma, \beta, \alpha)$  at  $t = 0$ ;
2. Compute  $\delta_t$  using the contraction mapping  $\delta_t = \delta_{t-1} + \ln s - \ln \tilde{s}(\delta_{t-1}, \sigma)$  until convergence;
3. Compute residual  $\xi(\theta_t)$  and obtain  $\hat{\theta}^{GMM} \equiv \arg \min_{\theta} \vartheta(\theta_t)^\top W_J \vartheta(\theta_t)$ ;
4. Replace  $\theta_{t-1}$  with  $\hat{\theta}^{GMM}$  and go back to step 2; repeat until GMM criterion is minimised.

### 2.2.3 Asymptotics

The asymptotics are not straightforward as in the  $N \rightarrow \infty$  case, as changing  $J$  will mechanically change the market shares of **all other goods**. Furthermore, competition should force prices down. These complications are discussed in Berry Linton Pakes (2004) as well as the review article by Akerberg et al. (2007).

## 2.3 Estimation with Market-Level Data: Supply Side

A similar GMM estimator can be used to solve for the supply side of the model, either separately from demand or jointly. We impose an additive parametric form on the marginal cost as a function of exogenous cost shifter(s)  $w_j$ ,

$$mc_j = \gamma w_j + \omega_j \quad \mathbb{E}(\omega_j | w_j) = 0 \quad (25)$$

Profit maximisation implies

$$\frac{\partial \pi_j}{\partial p_j} \equiv \frac{\partial [s_j(p_j - mc_j)]}{\partial p_j} = \frac{\partial s_j}{\partial p_j}(p_j - mc_j) + s_j = 0 \Leftrightarrow p_j = \gamma w_j + \omega_j - s_j \left( \frac{\partial s_j}{\partial p_j} \right)^{-1} \quad (26)$$

which can be rearranged to isolate  $\omega_j$ . Note that  $\omega_j$  will depend on  $\theta$  via market shares. From the orthogonality condition

$$\mathbb{E}(\omega_j(\gamma, \theta) | w_j) = 0 \text{ with sample counterpart } \varphi(\gamma, \theta) \equiv \frac{1}{J} \sum_j^J \omega_j(\gamma, \theta) w_j \quad (27)$$

generating additional moments that can be stacked with  $\boldsymbol{\vartheta}(\theta)$  as  $\boldsymbol{\vartheta}^*(\theta, \gamma) = \{\boldsymbol{\vartheta}(\theta), \varphi(\gamma, \theta)\}^\top$ ; this yields the GMM estimator

$$(\hat{\theta}, \hat{\gamma}) \in \arg \min_{\theta, \gamma} \boldsymbol{\vartheta}^*(\theta)^\top W_J \boldsymbol{\vartheta}^*(\theta) \quad (28)$$

which can be estimated either directly or by iteration depending on the demand side.

### 3 Application: Berry Levinsohn Pakes (1995)

Analysis of the market of new cars: US-year treated as single market. Since very few people buy a car each year, **standard logit model would estimate very large substitution toward the ‘no-buy/used-car’ outside good**. Slightly different specification from ours, as the authors specify  $\delta_j$  as a function of  $\alpha(y_i - p_j)$  – i.e. disposable income after purchase – instead of  $\alpha p_j$ .

The authors estimate a standard RMCL model with supply side moments; they draw income from Census distributions. Instruments: car’s own characteristics (taken as exogenous); characteristics of cars from the same maker; characteristics of cars from all makers.

## 4 Identification

### 4.1 Intuition

Three traditional sets of instruments used to identify the set of **linear parameters**  $(\alpha, \beta)$ :

1. **BLP Instruments**: characteristics of products sold by rival firms, which affect own prices;
2. **Marginal Cost Shifters**: Variables that shift  $p_j$  exclusively via  $mc_j$ , not via utility.
3. **Hausman instruments**: Price of the same product in other markets, under the assumption that cost shocks are common but demand shocks are independent.

Identification of **non-linear parameters**  $\boldsymbol{\sigma}$  is also of particular interest, as they are effectively responsible for cross-product utility interdependence and deviations from IIA. Intuitively,  $\boldsymbol{\sigma}$  can be identified via observed deviations from IIA in substitution patterns; alternatively, identification can be achieved from data on ‘second choices’ indicated in surveys.

Assuming a mixed logit functional form, the **‘log-sum’ result** implies a market share for first choice  $j$  (for a number of draws  $\Xi$ )

$$\tilde{s}_j(\boldsymbol{\delta}, \boldsymbol{\sigma}) = \frac{1}{\Xi} \sum_i^\Xi \frac{e^{d_j + \sigma_\nu \nu_i \mathbf{x}_j}}{\sum_{\ell \in \mathcal{J}} e^{d_\ell + \sigma_\nu \nu_i \mathbf{x}_\ell}} \quad (29)$$

and a market share for second choice  $k$ , in absence of  $j$ ,

$$\tilde{s}_k(\boldsymbol{\delta}, \boldsymbol{\sigma}) = \frac{1}{\Xi} \sum_i \frac{e^{d_k + \sigma \nu_i \mathbf{x}_k}}{\sum_{\ell \in \mathcal{J}_{\setminus j}} e^{d_\ell + \sigma \nu_i \mathbf{x}_\ell}} \quad (30)$$

whence it can be observed that the mismatch between predicted correlation of the two choice shares and actual correlation can provide information on  $\sigma$ .

In analogue fashion, one can use data on **repeated choices**: if a subset of the goods is unavailable for some periods, can observe forced substitution patterns and their relationship with good characteristics.

## 4.2 Nonparametric Identification - Market-Level Data

Berry and Haile (2014) prove that  $(\alpha, \beta, \boldsymbol{\sigma})$  are identified **nonparametrically** - i.e. their identification does not rely on the parametric assumptions made so far. The necessary requirement is the existence of a ‘**special regressor**’  $x$  that does not have a random coefficient.

Assume availability of market-level data on market shares, characteristics and prices for  $J$  goods. The specification is given by

$$u_{ij} = x_j^{(1)} + \beta_i \mathbf{x}_j^{(2)} - \alpha p_j + \xi_j + \varepsilon_{ij} \quad (31)$$

where  $x^{(1)}$  is the special regressor, whose coefficient is normalised to 1, and  $\mathbb{E}(\xi_j | x_j^{(1)}, \mathbf{x}_j^{(2)}) = 0$ .

Identification is proved in three steps:

### Step 1: Index

Given (31), define the index  $\delta_j = x_j + \xi_j$ , excluding  $p_j$  and  $\mathbf{x}^{(2)}$ .

### Step 2: Invert

Under the conditions in Berry (1994), indices  $\boldsymbol{\delta}$  can be isolated from the rest of the utility function as  $\boldsymbol{\delta} = d(\mathbf{s}, \mathbf{p})$ ; in turn, this can be rewritten as

$$x_j = d(\mathbf{s}, \mathbf{p}) - \xi_j \quad (32)$$

which has a separable error  $\xi_j$  and two sets of endogenous variables  $(\mathbf{s}, \mathbf{p})$ .

### Step 3: Instrument

Use the identification argument of Newey and Powell (2003) for nonparametric specifications of the form

$$y = f(\mathbf{x}) + \boldsymbol{\xi} \quad \mathbb{E}[\boldsymbol{\xi} | \mathbf{Z}] = 0 \quad (33)$$

for which Newey and Powell prove identification provided sufficient relevance of the instrument – i.e. provided that any two functions of  $\mathbf{x}$  can be distinguished by variation in  $\mathbf{Z}$  ( $\mathbf{Z}$  is ‘complete’ in  $\mathbf{x}$ ). Note that (32) is not exactly amenable to (33), as we have the exogenous regressor on the LHS; however, Berry and Haile (2014) show that Newey and Powell’s proof can be used to establish identification of  $d(\cdot)$  under analogous completeness and exclusion conditions.



### 4.3 Some Parametric Examples

#### 4.3.1 Logit

$$\xi_j = \ln s_j - \ln s_0 - \beta \mathbf{x}_j + \alpha p_j \Rightarrow x_j^{(1)} = \frac{1}{\beta^{(1)}}(\ln s_j - \ln s_0) - \frac{\beta^{(2)}}{\beta^{(1)}} \mathbf{x}_j^{(2)} + \frac{\alpha}{\beta^{(1)}} p_j - \frac{1}{\beta^{(1)}} \xi_j$$

need instruments for  $(\ln s_j - \ln s_0)$  and for  $p_j$ .

#### 4.3.2 Nested Logit

$$x_j^{(1)} = \frac{1}{\beta^{(1)}}(\ln s_j - \ln s_0) - \frac{\beta^{(2)}}{\beta^{(1)}} \mathbf{x}_j^{(2)} + \frac{\alpha}{\beta^{(1)}} p_j - \frac{1}{\beta^{(1)}} \xi_j - \frac{\sigma_r}{\beta^{(1)}} \ln s_{j \in \mathcal{J}_r}$$

now we need instruments for  $(\ln s_j - \ln s_0)$ , for  $p_j$ , **and** for  $\ln s_{j \in \mathcal{J}_r}$ . More general substitution patterns require availability of more instruments.

#### 4.3.3 Mixed Logit

$$u_{ij} = (\beta + \sigma_\nu \nu_i) \mathbf{x}_j - \alpha p_j + \xi_j + \varepsilon_{ij} \quad (34)$$

yields a market share function  $\mathbf{s} = \mathbf{s}(\delta, \mathbf{p}, \sigma)$  which can be inverted as  $x_j = d_j(\mathbf{s}, \mathbf{p}, \sigma) - \xi_j$ , which shows that now instruments for **vectors**  $(\mathbf{s}, \mathbf{p})$  are necessary;  $\mathbf{s}$  can be instrumented by  $\mathbf{x}$ .

#### 4.3.4 Summary

Although the literature has always highlighted the necessity of instrumenting for endogenous prices, Berry and Haile show that this is not enough – that **one also needs to instrument for market shares**. The construction of the three-step process provides an instrument for  $s_j$  in  $x_j$ .

### 4.4 Nonparametric Identification - Agent-Level Data

Berry and Haile also analyse the case in which one possesses data on product characteristics, product prices, purchase choices, and some variable  $z_{ijt}$  that varies at the (individual  $\times$  good  $\times$  market) level.

$$u_{ijt} = (z_{ijt} + \xi_{jt}) - \alpha p_{jt} + \varepsilon_{ijt} \quad (35)$$

where the coefficient on  $z_{ijt}$  is normalised to 1 and  $z_{ijt} + \xi_{jt} = \delta_{ijt}$ . From the micro data, one observes conditional choice probabilities as a function of observed  $\hat{\mathbf{z}}, \mathbf{s}_t(\hat{\mathbf{z}})$ ; in turn, for any share  $s$  generated by some  $\hat{\mathbf{z}}$  there exist at least one vector of characteristics  $z_t^*(s)$  such that  $s_t(z_t^*(s)) = s$ .

The identification argument in presence of micro data proceeds in **two broad steps**: first, variation in  $z_{it}$  is used to obtain a scalar variable that is a function of  $\xi_{jt}$  and prices; second, instrumental variables are used to identify parameters of the function and  $\xi_{jt}$ .

1. Fix a probability vector  $\mathbf{s}^*$  across markets;
2. Obtain  $\mathbf{z}_t(\mathbf{s}^*) \forall t$ ;
3. Invert market share function  $\mathbf{s}^* = \mathbf{s}(\delta_{it}, \mathbf{p}_t)$  to obtain  $\delta_{ijt} = d(\mathbf{s}^*, \mathbf{p}_t)$ ;
4. From  $z_{ijt} + \xi_{jt} = \delta_{ijt}$  we have  $\mathbf{z}_{jt}(\mathbf{s}^*) = d_t(\mathbf{s}^*, \mathbf{p}_t) - \xi_{jt}$ ;
5. Need to only instrument for  $\mathbf{p}$  as  $\mathbf{s}^*$  is fixed.<sup>3</sup>

---

<sup>3</sup>Instrumenting for market shares intuitively becomes redundant as variation in  $z_{it}$  will identify variation in quantities. See Berry and Haile (2016, Section 4.1) for further discussion.

## 5 Estimation with Agent-Level Data

Random Sample of  $N$  consumers indexed by  $i$  belonging to  $J$  markets indexed by  $j$ . Data on market characteristics, consumer characteristics, purchase decision  $y_{ij}$ .

Given  $u_{ij} = \delta_j + \sigma \nu_i \mathbf{x}_j + \varepsilon_{ij}$ , the model yields a  $P_{ij}(\theta, \sigma, \xi | \nu_i)$  as long as we specify a distribution for  $\varepsilon_{ij}$ . It is important to note that  $\xi$  and  $\theta = (\alpha, \beta)$  **cannot separately be identified**, as each  $P_{ij}$  is defined by a pair  $(\alpha, \beta)$  that deterministically implies a value of  $\xi$ ; we can at best identify  $\delta$ . As a result, we turn to a two-step method, the first step of which can be computed in two different ways. In the first step(s), we estimate  $(\delta, \sigma)$ ; in the second step we obtain  $\xi$  and estimate  $(\alpha, \beta)$ .

### Step 1.a – Maximum Likelihood

The probability that  $i$  chooses  $j$  is defined as

$$\Pr_j(\delta, \sigma, \nu_i) = \int_{\varepsilon} \mathbb{1}\{u_{ij} > u_{ik} \ \forall j \neq k\} dG(\varepsilon) \quad (36)$$

Hence the MLE is the  $(\delta, \sigma)$  vector that maximises the log-likelihood of observed market shares, i.e.

$$\ln \mathcal{L}(\delta, \sigma) = \sum_i^N \sum_j^J \mathbb{1}\{y_{ij} = 1\} \ln P_j(\delta, \sigma, \nu_i) \quad (37)$$

where, to speed up computation, one can separately obtain  $\delta$  as  $\delta(s, \sigma)$  via BLP contraction mapping.

### Step 1.b – Generalised Method of Moments

Alternatively, with instruments  $[Z_{ij}^1, \dots, Z_{ij}^M]$ ,  $(\delta, \sigma)$  solve (given just-identification)

$$\sum_i \sum_j [y_{ij} - P_j(\delta(s, \sigma), \sigma, \nu_i)] Z_i^m = 0 \ \forall m \quad (38)$$

The advantage of GMM over ML is that if the conditional probabilities  $P_{ij}(\cdot | \nu_i)$  are **simulated**, ML will lead to inconsistent estimates. In fact, albeit simulated probabilities are a consistent estimator of the real probabilities, logarithmic transformations of the simulated probabilities are **not consistent estimators of log-probabilities** (although bias will shrink as the number of draws increases).

### Step 2 – Estimation of Linear Parameters $(\alpha, \beta)$

Armed with  $\delta(s, \sigma)$  residuals are defined as

$$\xi(\alpha, \beta) = \delta - \beta \mathbf{x} + \alpha \mathbf{p}$$

and, for instruments  $[Z_{ij}^1, \dots, Z_{ij}^M]$ ,  $(\alpha, \beta)$  can be computed via linear GMM<sup>4</sup> using the instrument validity conditions as identifying restrictions:

$$\mathbb{E}[Z^m \xi(\beta_0, \alpha_0)] = 0$$

---

<sup>4</sup>Linear GMM as the specification of the residual is undoubtedly linear; choice does not depend on distributional assumptions on  $\varepsilon$ .

## 6 Application: Goolsbee and Petrin (2004)

### 6.1 Overview

- Focuses on direct broadcast satellites (DBS) as a competitor to cable; i.e. the paper is about how entry changes structure of competition and thus affects consumer welfare. Prior to DBS, cable faced essentially no competition in the US; after 1996, the industry was essentially completely de-regulated too, so had it not been for DBS, US cable companies would have enjoyed regional monopolies.
- Four options: Satellite (DBS); basic (BAS) and premium (PRE) cable; local antenna (ANT).
- Use micro data (30 000 households, 317 markets).
- Counterfactual analysis suggests that **without DBS entry cable prices would be about 15% higher and cable quality would fall**. Estimate a \$2.5 billion aggregate welfare gain for satellite buyers, and about \$3 billion for cable subscribers: **large welfare gains from availability of a new good and from increased competition**.
- **Strategy** in broad strokes: Estimate own- and cross-price elasticities for cable and satellite TV; model the supply side response of cable systems to the rise of DBS, examining how cable prices and characteristics respond to DBS; compute the implied change in consumer welfare caused by entry of DBS for adopters and non-adopters.

### 6.2 Demand System

- Significant cross-sectional variation in cable prices, no variation in satellite prices – hence rely on Slutsky symmetry to estimate price elasticities (implying impact of increasing price of DBS by one percent equals impact of decreasing price of substitutes by one percent):

$$s_{ANT} + s_{DBS} + s_{BAS} + s_{PRE} = 1 \Rightarrow \frac{\partial s_{ANT}}{\partial p_{DBS}} + \frac{\partial s_{DBS}}{\partial p_{DBS}} + \frac{\partial s_{BAS}}{\partial p_{DBS}} + \frac{\partial s_{PRE}}{\partial p_{DBS}} = 0$$

by Slutsky symmetry,  $\frac{\partial s_{DBS}}{\partial p_{DBS}} = -\frac{\partial s_{DBS}}{\partial p_{ANT}} - \frac{\partial s_{DBS}}{\partial p_{BAS}} - \frac{\partial s_{DBS}}{\partial p_{PRE}}$

- **Multinomial probit** – sensible since very few products, and allows for general patterns of substitution (elasticities unrestricted). With mixed logit, we're still assuming that the  $\epsilon_{ijt}$ 's are uncorrelated, whereas they can be correlated with the probit. Probit can allow e.g. for time-varying preferences for unobserved attributes of products (assuming they're normally distributed.)
- Include **separate product FE for each market** to control for unobserved product quality differences. Excluding these fixed effects yields demand elasticities strongly biased toward zero. Basic specification of the model:

$$U_{nj} = \delta_{mj} + p_{mj} \sum_{g=2}^5 \alpha_g \mathbb{1}_{\{n \in g\}} + \sum_{l=1}^L \beta_{jl} z_{nl} + \epsilon_{nj} \quad (39)$$

where  $n$  indexes individuals,  $j$  indexes products,  $l$  indexes individual characteristics,  $m$  indicates a market and  $g$  an income group. The first sum allows for **price coefficients to vary by income group**; the second corresponds to the **product fixed effects interacted with demographic**

**characteristics.** Note that this differs from mixed probit in that **we haven't interacted demographic variables with product characteristics.** The  $\epsilon_{mj}$ 's are normal and otherwise unrestricted. We also have

$$\delta_{mj} = \alpha_0 p_{mj} + \beta x_{mj} + \xi_{mj} \quad (40)$$

i.e. the usual product-level fixed effects which comprise of contributions from observed and unobserved product characteristics.

- For the second stage estimates (needed to get  $\alpha_0$  and  $\beta$ , both of which we need to compute elasticities) they try both SUR and 3SLS estimation (the first assumes prices are exogenous; the second allows them to be endogenous, and instruments for prices with local taxes.)
- **FINDINGS:** demand for DBS and for premium cable are more elastic than the demand for expanded basic; consumers view DBS to be a closer substitute for premium cable than it is for expanded basic. Strong covariance between  $\epsilon_{nj}$ 's is found (indicating logit would have been a bad choice). Instrument for price is important: SUR estimates are 1/6<sup>th</sup> of those for 3SLS estimates. Compensated and uncompensated demand curves are almost identical (little to no income effects.)

### 6.3 Supply Side Basics

- Model cable prices as a function of the observed and unobserved factors that affect the quality of DBS, cable, and premium television in the market and other exogenous factors such as average consumer demographics.
- Demand side estimates important here – provide a set of controls that reflect the unobserved quality and tastes for each product in each market.
- Differences in DBS quality are due to reception – require unobstructed line of sight to receive good coverage.
- Difference in cable characteristics include whether Pay Per View is available, channel capacity of the cable system, the price of basic plus expanded basic services, the price of premium services, the year the franchise system began (proxy for technology) and some other stuff.
- Results suggest higher DBS quality in a market is correlated with lower cable prices; evidence of modest improvements in cable quality in response to DBS entry. **Increased competition is welfare-improving for those who don't buy DBS as well.**
- **Reduced form pricing equation**, consisting partly of DBS and cable quality (the  $\xi$ 's estimated in the demand system). Supply side parameters are not obtained from first order pricing conditions: the authors argue this would require assuming equilibrium concept (but are we better off without that assumption?) and knowledge of marginal cost (no easy way around that one).

### 6.4 Estimation Basics

- Goolsbee and Petrin follow a two-step estimation procedure similar to BLP discussed in Section 5. They start by estimating the parameters  $\theta$ —these are the parameters that *are not* in the product fixed effects  $\delta$ , i.e. they are the group-specific price parameters and the demographic-fixed effect

interaction parameters—via maximum likelihood. Here the sample log-likelihood is

$$\ell(\boldsymbol{\theta}, \boldsymbol{\delta}) = \sum_{i=1}^N \sum_{j=0}^J y_{ij} \ln P_j(\boldsymbol{\delta}, \boldsymbol{\theta}, \mathbf{z}_i) \quad (41)$$

where  $y_{ij}$  is consumer  $i$ 's choice and  $\mathbf{z}_i$  are either observed individual characteristics (with micro data), simulated draws from your preferred distribution (with market-level data), or both (combined observed and unobserved taste heterogeneity). The key observation here is that the likelihood depends exclusively on  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$ , **not on components within  $\boldsymbol{\delta}$** , as other variables enter exclusively through  $\boldsymbol{\delta}$ .<sup>5</sup>

- To maximise the likelihood, we start with an initial guess of  $\boldsymbol{\theta}$  from which we obtain  $\boldsymbol{\delta}$ . The contraction mapping approach of BLP doesn't apply to probit, so Goolsbee and Petrin use nonlinear least squares (NLLS) to solve individually for each element in  $\boldsymbol{\delta}$ :

$$\delta_m(\boldsymbol{\theta}) = \arg \min_{\delta_m} \sum_j^3 (s_{mj}(\boldsymbol{\theta}, \delta_m) - \hat{s}_{mj}). \quad (42)$$

Since BLP show that the  $\delta_m(\boldsymbol{\theta})$  solving this equation exists and is unique, the unique minimiser of this equation for each  $m$  is  $\delta_m$  – we do this for all  $M$  markets, individually. Consistency comes from the size of the market, not the number of products – we're effectively assuming the market shares estimates from our data are observations on the true market shares—i.e. no estimation error.

We then maximise the likelihood with respect to  $\boldsymbol{\theta}$  treating  $\boldsymbol{\delta}$  fixed. From our new value of  $\boldsymbol{\theta}$  we then obtain a new value of  $\boldsymbol{\delta}$  from NLLS. We repeat this process until we reach some tolerance level for  $\boldsymbol{\theta}$ . This step in the procedure could be called *concentrated maximum likelihood* – we're concentrating out  $\boldsymbol{\delta}$  to estimate  $\boldsymbol{\theta}$ , and *vice versa*.

- Once this process is done we can proceed to estimate the rest of our model as usual: Via the equation  $\delta_j = \mathbf{x}'_j \boldsymbol{\beta} - \alpha p_j + \xi_j$ , using instruments such that  $E(\xi_j | \mathbf{x}_j, \mathbf{z}_j) = 0$ .

---

<sup>5</sup>A similar insight comes from thinking about the BLP two-step procedure outlined in Akerberg, Benkard, Berry and Pakes (2007).