Structural Estimation

Author: Wenxiao Yang

Institute: Haas School of Business, University of California Berkeley

Date: 2024

All models are wrong, but some are useful.

Contents

Chapter 1 Homogenous Products

How do we estimate demand for a homogenous product?

1.1 [Working(1927)]: OLS is not informative with endogeneity

Demand and Supply for coffee:

$$Q_t^d = \alpha_0 + \alpha_1 P_t + U_t$$

$$Q_t^s = \beta_0 + \beta_1 P_t + V_t$$

where $\alpha_1 < 0$ and $\beta_1 > 0$. Equilibrium price and quantity are given by

$$Q_t^d = Q_t^s \Rightarrow \begin{cases} P_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{V_t - U_t}{\alpha_1 - \beta_1} \\ Q_t = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 V_t - \beta_1 U_t}{\alpha_1 - \beta_1} \end{cases}$$

Note *Price is a function of both error terms, and we can't use a clever substitution to cancel things out.*

We can find that price is positively correlated with demand shift U_t and negatively correlated with supply shift V_t.

Consider the OLS estimator: $\hat{\alpha}_1 = \hat{\beta}_1 = \frac{\text{Cov}(P_t, Q_t)}{\text{Var}(P_t)}$. Since we have

$$\left.\begin{aligned} \text{Cov}(P_t, Q_t) = \alpha_1 \text{Var}(P_t) + \text{Cov}(P_t, U_t) \\ \text{Cov}(P_t, Q_t) = \beta_1 \text{Var}(P_t) + \text{Cov}(P_t, V_t) \end{aligned}\right\} \Rightarrow \begin{cases} \text{Bias}(\alpha_1) = |\hat{\alpha}_1 - \alpha_1| = \frac{\text{Cov}(P_t, U_t)}{\text{Var}(P_t)} \\ \text{Bias}(\beta_1) = |\hat{\beta}_1 - \beta_1| = \frac{\text{Cov}(P_t, V_t)}{\text{Var}(P_t)} \end{cases}$$

When $\text{Cov}(U_t, V_t) = 0$, the OLS estimator can be written as

$$\hat{\alpha}_1 = \hat{\beta}_1 = \frac{\alpha_1 \text{Var}(V_t) + \beta_1 \text{Var}(U_t)}{\text{Var}(V_t) + \text{Var}(U_t)}$$

More variation in supply V_t gives a better estimate of demand and more variation in demand U_t gives a better estimate of supply.

The OLS is not informative about the economic demand function (or supply function).

To deal with this problem, we need an excluded instrument that shifts one curve without affecting the other. Then, we can use this to form a 2SLS estimate.

1.2 [Angrist, Graddy, and Imbens(2000)]: IV is limited

1.2.1 Motivation

What if we don't generality know which kind of heterogeneity we face?

There are four cases that are ranked in increasing complexity:

1. Linear system with constant coefficients:
$$q_t^d(p,z,x) = \alpha_0 + \alpha_1 p + \alpha_2 z + \alpha_3 x + \epsilon_t$$
$$q_t^s(p,z,x) = \beta_0 + \beta_1 p + \beta_2 z + \beta_3 x + \eta_t$$
2. Linear system with non-constant coefficients:
$$q_t^d(p,z,x) = \alpha_{0t} + \alpha_{1t} p + \alpha_{2t} z + \alpha_{3t} x + \epsilon_t$$
$$q_t^s(p,z,x) = \beta_{0t} + \beta_{1t} p + \beta_{2t} z + \beta_{3t} x + \eta_t$$
3. Nonlinear system with constant shape (separable):
$$q_t^d(p,z,x) = q^d(p,z,x) + \epsilon_t$$
$$q_t^s(p,z,x) = q^s(p,z,x) + \eta_t$$
4. Nonlinear system with time-varying shape (non-separable): any forms of $q_t^d(p,z,x)$ and $q_t^s(p,z,x)$.

1.2.2 Model

We assume the regularity conditions (existence of first and second moment and being stationary) and $q_t^d(p,z,x)$, $q_t^s(p,z,x)$ are continuously differentiable in p.

Instrumental Variable Assume binary instrument $z_t \in \{0,1\}$ to make things easier. And $z_t \in \{0,1\}$ is a valid instrument in q_t^d, i.e., it satisfies

1. Exclusion: $q_t^d(p_t, z=1, x_t) = q_t^d(p_t, z=0, x_t) = q_t^d(p_t, x_t)$.
2. Relevance: $q_t^s(p_t, z=1, x_t) \neq q_t^s(p_t, z=0, x_t)$ for some period t.
3. Independence: ϵ_t, η_t, z_t are mutually independent conditional on x_t.

Suppose $z=1$ denote "stormy at sea" and $z=0$ denote "calm at sea". (Offshore weather makes fishing more difficult but doesn't change onshore demand.)

2SLS can work in linear models,
$$\hat{\alpha}_{1,0} = \frac{\mathbb{E}_t[\widehat{q_t|z_t=1}] - \mathbb{E}_t[\widehat{q_t|z_t=0}]}{\mathbb{E}_t[\widehat{p_t|z_t=1}] - \mathbb{E}_t[\widehat{p_t|z_t=0}]} \xrightarrow{P} \frac{\mathbb{E}_t[q_t|z_t=1] - \mathbb{E}_t[q_t|z_t=0]}{\mathbb{E}_t[p_t|z_t=1] - \mathbb{E}_t[p_t|z_t=0]} := \alpha_{1,0}$$
but it is not an estimator of a structural parameter in nonlinear models.

Claim 1.1

Authors make the point that IV estimator identifies something about relationship between p and q, without identifying deep structural parameters.

Some assumptions are needed to interpret the IV estimator.

Assumption

1. *Observed price is market clearing price* $q_t^d(p_t) = q_t^s(p_t, z_t)$ *for all* t*. (i.e., no friction).*
2. *For each value of* z *and* t*, there is a unique market clearing price,* $\tilde{p}(z,t)$*, such that*

$$q_t^d(\tilde{p}(z,t)) = q_t^s(\tilde{p}(z,t), z)$$

   $\tilde{p}(z,t)$ *is the potential price under any counterfactual* (z,t)*.*
3. $\mathbb{E}_t[p_t|z_t = 1] \neq \mathbb{E}_t[p_t|z_t = 0]$
4. $\tilde{p}(z,t)$ *is weakly increasing in* z*.*

Lemma 1.1

The numerator of $\alpha_{1,0}$ can be given by

$$\mathbb{E}_t[q_t|z_t = 1] - \mathbb{E}_t[q_t|z_t = 0] = \mathbb{E}_t\left[\int_{\tilde{p}(0,t)}^{\tilde{p}(1,t)} \frac{\partial q_t^d(s)}{\partial s} ds\right]$$

Theorem 1.1

Based on this lemma, the IV estimator equals to

$$\begin{aligned}\alpha_{1,0} &= \frac{\mathbb{E}_t\left[\int_{\tilde{p}(0,t)}^{\tilde{p}(1,t)} \frac{\partial q_t^d(s)}{\partial s} ds\right]}{\mathbb{E}_t\tilde{p}(1,t) - \mathbb{E}_t\tilde{p}(0,t)} \\ &\to \int_0^\infty \mathbb{E}_t\left[\frac{\partial q_t^d(s)}{\partial s} \mid s \in [\tilde{p}(0,t), \tilde{p}(1,t)]\right] \omega(s) ds\end{aligned}$$

where $\omega(s)$ is the weight that is not a function of t, but it is largest for prices most likely to fall in $[\tilde{p}(0,t), \tilde{p}(1,t)]$.

Note

1. $\alpha_{1,0}$ *only provides information about demand curve in range of potential price variation induced by the instrument.*
2. *For different instruments* z*,* $\alpha_{1,0}$ *has a different interpretation like the LATE does. (Different from the linear model where anything works!).*

Chapter 2 Random Utility Models

2.1 Gumbel Distribution

Definition 2.1 (Gumbel Distribution)

Gumbel distribution is also called **type-I generalized extreme value distribution**, denoted by Gumbel(μ, β). The μ is the location and $\beta > 0$ is the scale.

Its cdf is

$$F(x, \mu, \beta) = e^{-e^{-(x-\mu)/\beta}},$$

and its pdf is

$$f(x, \mu, \beta) = \frac{1}{\beta} e^{-(z+e^{-z})}, \text{ where } z = \frac{x-\mu}{\beta}$$

Lemma 2.1 (Properties of Gumbel Distribution)

The properties of Gumbel distribution are given as follows.

1. The mean is $\mathbb{E}[X] = \mu + \gamma\beta$, where γ is the Euler-Mascheroni constant.
2. The median is $\mu - \beta \ln(\ln 2)$;
3. The variance is $\frac{\pi^2}{6}\beta^2$;
4. If $G_1, ..., G_k$ are i.i.d. Gumbel random variables with parameters (μ, β), its maximum is also a Gumbel random variable,

$$\max\{G_1, ..., G_k\} \sim \text{Gumbel}(\mu + \beta \ln k, \beta).$$

   Then, $\mathbb{E}[\max\{G_1, ..., G_k\}] = \mu + \beta \ln k + \gamma\beta$.

Lemma 2.2

Suppose $g_1, ..., g_n$ be independent samples of Gumbel$(0, 1)$. Then,

1. $\text{argmax}_i(g_i + \delta_i) \sim \text{Categorical}\left(\frac{e^{\delta_j}}{\sum_i e^{\delta_i}}\right)_j$.
2. $\max_i(g_i + \delta_i) \sim \text{Gumbel}\left(\log\left(\sum_i e^{\delta_i}\right), 1\right)$.
3. $\mathbb{E}[\max_i(g_i + \delta_i)] = \log\left(\sum_i e^{\delta_i}\right) + \gamma$.

2.2 Random Utility Models

Content is based on [Berry and Haile(2021)].

Let $j = 1, \ldots, J_i$ index the "inside goods" available to consumer i while $j = 0$ denotes the outside good. A consumer's choice set is characterized by J_i and a set χ_i, which may include

- observed characteristics of consumer i,
- observed characteristics of goods (including prices),
- observed characteristics of the local market,
- and characteristics of the market or goods that are unobserved to the researcher.

Each consumer i has a (conditional indirect) utility u_{ij} for good j. Consumer knows her utilities for all goods and chooses the good with the highest utility.

In this model, the heterogeneity of consumer preferences is modeled by random utilities:

Definition 2.2 (Random Utility Model)

Given the choice set (J_i, χ_i), each consumer's utility vector $(u_{ij})_{j=0,1,\ldots,J_i}$ is an independent draw from a joint distribution $F_u(\cdot \mid J_i, \chi_i)$.

Since only the ordinal ranking of goods matters for a consumer's behavior, we can normalize the location and scale of each consumer's utility vector without loss of generality. We assume that "ties" ($u_{ij} = u_{ik}$ for some $j \neq k$) occur with probability zero in the distribution $F_u(\cdot \mid J_i, \chi_i)$. Then, we can represent consumer i's choice with the vector $(q_{i1}, \ldots, q_{iJ_i})$, where

$$q_{ij} = \mathbf{1}\{u_{ij} \geq u_{ik}, \forall k\}$$

The consumer-specific choice probabilities are then given by

$$s_{ij} := \mathbb{E}[q_{ij} \mid J_i, \chi_i] = \int_{\mathcal{A}_{ij}} dF_u\left(u_{i0}, \ldots, u_{iJ_i} \mid J_i, \chi_i\right)$$

where $\mathcal{A}_{ij} = \{(u_{i0}, \ldots, u_{iJ_i}) \in \mathbb{R}^{J_i+1} \mid u_{ij} \geq u_{ik}, \forall k\}$.

2.3 The Canonical Model

Definition 2.3 (Canonical Model: Independence of Irrelevant Alternatives (IIA))

Discrete choice demand models are frequently formulated using a parametric random utility specification such as

$$u_{ijt} = x_{jt}\beta_{it} - \alpha_{it}p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

for $j > 0$, with $u_{i0t} = \epsilon_{i0t}$.

The notion of a "market" t allows a precise characterization of the endogeneity problems inherent to demand estimation (In practice, markets are typically defined by natural combinations of time and geography). Let $\mathcal{J}_t$ denote the set of products (inside goods) available to consumers in market t, and let $J_t = |\mathcal{J}_t|$.

Let $x_t = (x_{1t}, ..., x_{J_t,t}), p_t = (p_{1t}, ..., p_{J_t,t}), \xi_t = (\xi_{1t}, ..., \xi_{J_t,t})$, and $\chi_t = (x_t, p_t, \xi_t)$.

1. p_{jt} represents the price of good j in market t, while $x_{jt} \in \mathbb{R}^K$ represents other observable characteristics of good j in the market.
2. ξ_{jt} is *a demand shock*, an unobserved factor associated with good j and market t.
   ◦ (ξ_{jt} can represent any combination of latent taste variation and latent product characteristics common to consumers in market t. For example, a high value of ξ_{jt} may simply indicate that consumers in market t have a high mean taste for good j.)
   ◦ (ξ_{jt} is correlated with p_{jt} and x_{jt} by the endogeneity of prices and additional characteristics.)
   ◦ (ξ_{jt} is not a characteristic, i.e., $\mathbb{E}[\xi_{jt} \mid x_{jt}] = 0$.)
3. ϵ_{ijt} is the *utility shock*. It is most often specified as an i.i.d. draw from a standard type-1 extreme value distribution, yielding a *mixed multinomial logit model*.
   ◦ Choice probabilities in the population reflect a mixture of the choice probabilities conditional on each possible combination of $(\alpha_{it}, \beta_{it})$. In this case, the choice probabilities in the population (i.e., the market shares) are given by
$$s_{jt} = \int \frac{\exp(x_{jt}\beta_{it} - \alpha_{it}p_{jt} + \xi_{jt})}{\sum_{k=0}^{J_t} \exp(x_{kt}\beta_{it} - \alpha_{it}p_{kt} + \xi_{kt})} dF(\alpha_{it}, \beta_{it}; t) \tag{2.1}$$
   where the latent taste parameters α_{it} and β_{it} are often referred to as "random coefficients," and $F(\cdot; t)$ denotes their joint distribution in market t.
   ◦ Alternatively, a normal distribution will yield a mixed multinomial probit.
4. The joint distribution $F(\cdot; t)$ is commonly specified as follows.
   (a). Each component k of the random coefficient vector β_{it} typically specified takes the form
$$\beta_{it}^{(k)} = \beta_0^{(k)} + \beta_v^{(k)} v_{it}^{(k)} + \sum_{l=1}^{L} \beta_d^{(l,k)} d_{ilt} \tag{2.2}$$
   Remind that the $\beta_{it}^{(k)}$ contributes as $x_{jt}^{(k)}\beta_{it}^{(k)}$.
      I. The $\beta_0^{(k)}$ is a parameter shifting <u>all</u> consumers' tastes for additional characteristics $x_{jt}^{(k)}$.
      II. Each d_{ilt} represents a *characteristic* (e.g., demographic measure) of individual i; the $\beta_d^{(l,k)}$ governs the extent of variation in tastes for $x_{jt}^{(k)}$ with different values of d_{ilt}.
      III. Each $v_{it}^{(k)}$ represents a *taste shock*, which is a random variable with a pre-specified distribution

(e.g., a standard normal); the $\beta_v^{(k)}$ governs the extent of variation in tastes for $x_{jt}^{(k)}$ with different values of $v_{it}^{(k)}$.

(b). A typical specification of α_{it} takes the form

$$\ln(\alpha_{it}) = \alpha_0 + \alpha_y y_{it} + \alpha_v v_{it}^{(0)} \tag{2.3}$$

where y_{it} represents consumer-specific measures such as income that are posited to affect price sensitivity. The variables included in y_{it} might overlap partially or entirely with d_{it}.

2.4 Market-Level Data

We typically observe key data at the market level:

1. J_t: the number of goods available to consumers in each market t;
2. p_t, x_t: prices and additional characteristics of goods in each market t;
3. $\tilde{s}_{jt}$: observed market shares, typically measured by $\tilde{s}_{jt} := \frac{\text{total quantity of good } j \text{ sold in market } t}{\text{total number of consumers in market } t}$.
4. *Distributions of* (d_{it}, y_{it}): distribution of consumer characteristics in each market.
5. Additional variables w_t that might serve as appropriate instruments.

2.5 [McFadden(1972), McFadden(1981)]

2.5.1 Elasticity of Multinomial Logit Model

Consider the multinomial logit model with $u_{ij} = \beta x_j + \epsilon_{ij}$, the market shares are given by

$$s_{ij} = \frac{\exp(\beta x_j)}{\sum_{k=0}^{J} \exp(\beta x_k)}$$

where

$$\frac{\partial s_{ij}}{\partial x_j} = \beta s_{ij}(1 - s_{ij}); \quad \frac{\partial s_{ij}}{\partial x_k} = -\beta s_{ij} s_{ik}$$

Then, the **elasticity** and **cross elasticity** are given by

$$\frac{\partial \log s_{ij}}{\partial \log x_j} = \beta x_j(1 - s_{ij}); \quad \frac{\partial \log s_{ij}}{\partial \log x_k} = -\beta x_k s_{ik}$$

By the Lemma 2.2, the **expected utility of the consumer** i is

$$\mathbb{E}[\max_{j \in \{0,1\ldots,J\}} u_{ij}] = \log\left(\sum_{k=0}^{J} \exp(\beta x_k)\right) + \gamma$$

The **diversion ratio** is the probability that consumer i, who leaves product j after a decrease in x_j, switches to product k:

$$D_{ijk} = \left| \frac{\frac{\partial s_{ik}}{\partial x_j}}{\frac{\partial s_{ij}}{\partial x_j}} \right| = \frac{s_{ik}}{1 - s_{ij}}$$

2.5.2 Nested Logit Model

A traditional relaxation of IIA is the Nested Logit Model, which is usually presented as two sequential decisions: The utility of i from product j is given by

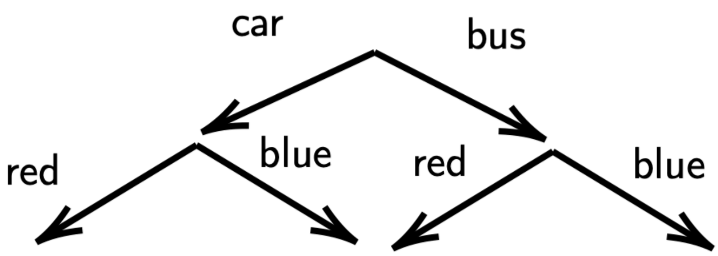

Figure 2.1: Nested Logit Model

$$u_{ij} = \delta_{ij} + \zeta_{ig} + (1 - \rho)\epsilon_{ij}$$

where δ_{ij} is the utility we want to estimate in multinomial logit, ζ_{ig} is a common preferences over all products that belong to the nest g.

ρ is often interpreted as a within-nest correlation of preferences. If $\rho = 0$, we are in the simple logit case.

Note *IIA still holds within each nest and also between the nests.*

The distribution of $\zeta_{ig} + (1 - \rho)\epsilon_{ij}$ is a special case of a GEV distribution such that

$$p_{ij} = \frac{\exp\left(\frac{\delta_{ij}}{1-\rho}\right)\left(\sum_{k\in g} \exp\left(\frac{\delta_{ik}}{1-\rho}\right)\right)^{-\rho}}{\sum_{h\in G}\left(\sum_{k\in h} \exp\left(\frac{\delta_{ik}}{1-\rho}\right)\right)^{1-\rho}}$$

where G is the set of nests. This probability can be decomposed into two probabilities:

$$p_{ij} = \underbrace{\frac{\exp\left(\frac{\delta_{ij}}{1-\rho}\right)}{\sum_{k\in g} \exp\left(\frac{\delta_{ik}}{1-\rho}\right)}}_{p_{ij|g}} \underbrace{\frac{\exp\left((1-\rho)\mathbb{E}[\max_{j\in g} u_{ij}]\right)}{\sum_{h\in G} \exp\left((1-\rho)\mathbb{E}[\max_{j\in h} u_{ij}]\right)}}_{p_{ig}}$$

where $\mathbb{E}[\max_{j\in h} u_{ij}] = \log\left(\sum_{k\in h} \exp\left(\frac{\delta_{ik}}{1-\rho}\right)\right) + \gamma$.

2.6 [Berry(1994)]: IV-based estimation with unobserved demand shock

Consider the case without random coefficients, i.e., (α, β) are common across all i:

$$u_{ijt} = \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{:=\delta_{jt}} + \epsilon_{ijt}$$

In MNL case, the choice probabilities in the population are given by

$$s_{jt} = \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J_t} \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})} = \frac{\exp(\delta_{jt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt})} := \hat{s}_{jt}(\delta_{1t}, ..., \delta_{J_t t}) \tag{2.4}$$

for $j = 1, ..., J_t$ and $s_{0t} = \frac{1}{1+\sum_{k=1}^{J_t} \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})}$.

Note *We cannot do **nonlinear least squares, i.e.***

$$\min_{\alpha,\beta} \sum_{j=1}^{J_t} \left(\tilde{s}_{jt} - \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J_t} \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})} \right)^2$$

because we need to know the x_{jt} in order to estimate α and β.

2.6.1 Welfare

By the Lemma 2.2, the consumer's welfare based on utility is given as

$$\mathbb{E}[\max_j u_{ijt}] = \log \left(1 + \sum_{j=1}^{J} \exp(\delta_{jt}) \right) + \gamma.$$

Suppose the δ_{jt} change to δ'_{jt}. Then the welfare change in utility is

$$\log \left(1 + \sum_{j=1}^{J} \exp(\delta'_{jt}) \right) - \log \left(1 + \sum_{j=1}^{J} \exp(\delta_{jt}) \right),$$

and the welfare change in dollars is

$$\Delta CW = \frac{1}{\alpha} \left[\log \left(1 + \sum_{j=1}^{J} \exp(\delta'_{jt}) \right) - \log \left(1 + \sum_{j=1}^{J} \exp(\delta_{jt}) \right) \right].$$

2.6.2 Inversion

[Berry(1994)] suggests an IV-based estimation with unobserved demand shock:

Assume there exist instruments z_{jt} such that $\mathbb{E}[\xi_{jt} z_{jt}] = 0$. The corresponding moment condition is

$$\frac{1}{J_t} \sum_{j=1}^{J_t} \xi_{jt} z_{jt} = \frac{1}{J_t} \sum_{j=1}^{J_t} (\delta_{jt} - x_{jt}\beta + \alpha p_{jt})\, z_{jt} \xrightarrow{J \to \infty} 0 \tag{Sample Moment}$$

which converges to zero at the true value of α and β. We want to estimate α and β by minimizing the sample moment. However, we do not know δ_{jt}.

Definition 2.4

[Berry(1994)] suggests a two-step approach

1. **Inversion**: By equating the data $\tilde{s}_{jt}$ and the choice probabilities $\hat{s}_{jt}(\delta_{1t}, ..., \delta_{Jt})$, we have a

system of J nonlinear equations:

$$\begin{cases} \tilde{s}_{1t} & = \hat{s}_{1t}(\delta_{1t}, ..., \delta_{J_t t}) \\ \vdots & \quad \vdots \\ \tilde{s}_{J_t t} & = \hat{s}_{J_t t}(\delta_{1t}, ..., \delta_{J_t t}) \end{cases}$$

Then, we can ``inverse'' this system of equations to solve for $\delta_{1t}, ..., \delta_{J_t t}$ as a function of $\tilde{s}_{1t}, ..., \tilde{s}_{J_t t}$:

$$\hat{\delta}_{jt} := \delta_{jt}(\tilde{s}_{1t}, ..., \tilde{s}_{J_t t}) \tag{2.5}$$

2. **IV Estimation**: Now, by going back to definition of δ_{jt}, we have

$$\begin{cases} \delta_{1t}: & = x_{1t}\beta - \alpha p_{1t} + \xi_{1t} \\ \vdots & \quad \vdots \\ \delta_{J_t t}: & = x_{J_t t}\beta - \alpha p_{J_t t} + \xi_{J_t t} \end{cases}$$

Now, using estimated $\hat{\delta}_{jt}$ to calculate sample moment, and find α and β by minimizing the sample moment (Sample Moment):

$$\min_{\alpha,\beta} \frac{1}{J_t} \sum_{j=1}^{J_t} \left(\hat{\delta}_{jt} - x_{jt}\beta + \alpha p_{jt} \right) z_{jt}$$

(If δ_{jt} is linear, we can use linear IV estimation.)

Example 2.1 (MNL Case)

Consider the MNL case in (2.4): Taking logs

$$\ln s_{0t} = -\log \left(1 + \sum_{k=1}^{J_t} \exp(\delta_{kt}) \right)$$

$$\ln s_{jt} = \delta_{jt} - \log \left(1 + \sum_{k=1}^{J_t} \exp(\delta_{kt}) \right)$$

Then, the equation system gives:

$$\underbrace{\ln \tilde{s}_{jt} - \ln \tilde{s}_{0t}}_{\text{Data}} = \delta_{jt} := x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

There is one to one mapping between s_{jt} and ξ_{jt}.

1. **Pro:** Now we can do *IV regression* of $\ln \tilde{s}_{jt} - \ln \tilde{s}_{0t}$ on $x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ with IV z_{jt}. 2SLS: regress p_{jt} on z_{jt} first and then regress $\ln \tilde{s}_{jt} - \ln \tilde{s}_{0t}$ on x_{jt} and $\hat{p_{jt}}(z_{jt})$.
2. **Con:** we need aggregate data and shares *without sampling error (variance)*. Note that $\ln(s_1 + \epsilon) - \ln(s_2 + \epsilon) \neq \ln(s_1) - \ln(s_2)$.

Note *We need* $\mathbb{E}[\xi_{jt} \mid x_{jt}] = 0$.

2.6.3 Elasticity

The logit model simplifies estimation but doesn't solve the IIA issues:

$$\frac{\partial s_{ij}}{\partial p_j} = -\alpha s_{ij}(1-s_{ij}); \quad \frac{\partial s_{ij}}{\partial p_k} = \alpha s_{ij}s_{ik}$$
$$\frac{\partial \log s_{ij}}{\partial \log p_j} = -\alpha p_j(1-s_{ij}); \quad \frac{\partial \log s_{ij}}{\partial \log p_k} = \alpha p_k s_{ik}$$

2.6.4 Supply Side Recover

In previous section, we have estimated the demand function for brand j, which is denoted by

$$s_j\left(\vec{x}, \vec{p}, \vec{\xi}\right)$$

where $\vec{x} := x_t = (x_{1t}, ..., x_{J_t,t}), \vec{p} := p_t = (p_{1t}, ..., p_{J_t,t}), \vec{\xi} := \xi_t = (\xi_{1t}, ..., \xi_{J_t,t})$. We omit the t in the notation. Now, we specify the costs of producing brand j as

$$C^j\left(q_j, w_j, \omega_j\right)$$

where q_j is total production of brand j, w_j are observed cost components associated with brand j (e.g. could be characteristics of brand j), ω_j are unobserved cost components (another structural error)

Then profits for brand j are:

$$\Pi_j = s_j\left(\vec{x}, \vec{p}, \vec{\xi}\right)p_j - C^j\left(s_j\left(\vec{x}, \vec{p}, \vec{\xi}\right), w_j, \omega_j\right)$$

For multiproduct firm: assume that firm k produces all brands $j \in \mathcal{K}$. Then its profits are

$$\tilde{\Pi}_k = \sum_{j\in\mathcal{K}} \Pi_j$$

The most common assumption is Bertrand (price) competition. Under price competition, equilibrium prices are characterized by J equations

$$\frac{\partial \tilde{\Pi}_k}{\partial p_j} = s_j + \sum_{j'\in\mathcal{K}} \frac{\partial s_{j'}}{\partial p_j}\left(p_j - \frac{\partial C^{j'}}{\partial q_{j'}}\right) = 0, \ \forall j \in \mathcal{K}, \forall k$$

where $\frac{\partial C^{j'}}{\partial q_{j'}}$ is the marginal cost function.

Since we have already estimated the demand side, all s_j and $\frac{\partial s_{j'}}{\partial p_j}$ can be calculated. Hence, we can solve the marginal costs $\frac{\partial C^j}{\partial q_j}$ as

$$\left(\frac{\partial C^1}{\partial q_1}, ..., \frac{\partial C^J}{\partial q_J}\right) = \vec{p} + (\Delta s)^{-1}\vec{s}$$

where $\vec{s} := (s_1, ..., s_J)$ and Δs is a $J \times J$ matrix where

$$\Delta s_{(i,j)} = \begin{cases} \frac{\partial s_i}{\partial p_j} & \text{if models } (i,j) \text{ produced by the same firm} \\ 0 & \text{otherwise} \end{cases}$$

Algorithms to Recover Equilibrium Prices Sometimes we want to recover the equilibrium prices in a counterfactual setting. Note that any equilibrium prices should satisfy

$$\vec{p} = \vec{C} - (\Delta s)^{-1}\vec{s}(\vec{p}).$$

By Newton's method, we can recover the equilibrium prices by iteration:

1. Initial guess: $\vec{p}^{(0)}$
2. Repeat $\vec{p}^{(k+1)} = \vec{C} - (\Delta s)^{-1}\vec{s}(\vec{p}^{(k)})$ until $\|\vec{p}^{(k+1)} - \vec{p}^{(k)}\| < 10^{-5}$.

Example 2.2

The simplest single-product firm problem

$$\max_{p_j} (p_j - c_j) \cdot s_j(p_j, p_{-j})$$

The equilibrium prices are given by

$$p_j^* = c_j - \frac{s_j}{\frac{\partial s_j}{\partial p_j}} = c_j + \frac{1}{\alpha(1-s_j)}$$

2.6.5 Instruments

1. Excluded cost shifters:

   Things that affect marginal cost but do not affect demand (e.g., ingredients prices, shipping costs)

2. Hausman instruments:

   Price of same good in another market. The idea is that marginal costs across markets are correlated but demand shocks ξ_{jt} are not.

3. Waldfogel instruments:

   Characteristics of nearby markets. Firms may set the same price for all cities within a region. The age/income/education in San Francisco may affect prices in Berkeley but not Berkeley preferences.

4. BLP instruments:

   Other product characteristics x_{-jt} are excluded from the estimating equation but affect prices due to changes in competition

2.6.6 Nested Logit Model

An analogous inversion that can be done for the nested logit model

$$\log(s_{jt}) - \log(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + \rho \log\left(s_{jt|g}\right) + \xi_{jt}$$

where $s_{jt|g}$ is the within-nest market share of product j.

$$s_{ij} = \frac{\exp\left(\frac{\delta_j}{1-\rho}\right)\left(\sum_{k\in g}\exp\left(\frac{\delta_k}{1-\rho}\right)\right)^{-\rho}}{\sum_{h\in G}\left(\sum_{k\in h}\exp\left(\frac{\delta_k}{1-\rho}\right)\right)^{1-\rho}} = \underbrace{\frac{\exp\left(\frac{\delta_j}{1-\rho}\right)}{\sum_{k\in g}\exp\left(\frac{\delta_k}{1-\rho}\right)}}_{:=s_{j|g}} \underbrace{\frac{\left(\sum_{k\in g}\exp\left(\frac{\delta_k}{1-\rho}\right)\right)^{1-\rho}}{\sum_{h\in G}\left(\sum_{k\in h}\exp\left(\frac{\delta_k}{1-\rho}\right)\right)^{1-\rho}}}_{:=s_g}$$

We have two endogenous variables p_{jt} and $\log\left(s_{jt|g}\right)$. So, in order to identify ρ, we need an additional instrument that moves the within-nest market share but is not correlated with ξ_{jt}, e.g.,

1. The number of other products in my nest
2. The characteristics of other products in my nest
3. Cost-shifters of other products in my nest

(First two instruments rely on product availability/characteristics to be random across markets.)

Derivatives and Elasticities The derivatives are given by

$$\frac{\partial s_g}{\partial \delta_k} = \begin{cases} s_k(1-s_g), & k\in g \\ -s_g s_k, & k\notin g \end{cases}$$

$$\frac{\partial s_{j|g}}{\partial \delta_k} = \begin{cases} \frac{1}{1-\rho}s_{j|g}(1-s_{j|g}), & k=j \\ -\frac{1}{1-\rho}s_{j|g}s_{k|g}, & k\neq j, k\in g \\ 0, & k\notin g \end{cases}$$

$$\begin{aligned}\frac{\partial s_j}{\partial \delta_k} &= \frac{\partial s_{j|g}}{\partial \delta_k}s_g + \frac{\partial s_g}{\partial \delta_k}s_{j|g} \\ &= \begin{cases} \frac{1}{1-\rho}s_j\left(1-\rho s_{j|g}-(1-\rho)s_j\right), & k=j \\ -s_k\left(s_j+\frac{\rho}{1-\rho}s_{j|g}\right) = -s_k s_j\left(1+\frac{\rho}{1-\rho}\frac{1}{s_g}\right), & k\neq j, k\in g \\ -s_j s_k, & k\notin g \end{cases}\end{aligned}$$

The price elasticities are given by

$$\frac{\partial \log s_j}{\partial \log p_k} = \frac{\frac{\partial \log s_j}{\partial s_j}}{\frac{\partial \log p_k}{\partial p_k}}\frac{\partial s_j}{\partial \delta_k}\frac{\partial \delta_k}{\partial p_k} = -\alpha\frac{p_k}{s_j}\frac{\partial s_j}{\partial \delta_k} = \begin{cases} -\frac{\alpha}{1-\rho}p_j\left(1-\rho s_{j|g}-(1-\rho)s_j\right), & k=j \\ \alpha p_k s_k\left(1+\frac{\rho}{1-\rho}\frac{1}{s_g}\right), & k\neq j, k\in g \\ \alpha p_k s_k, & k\notin g \end{cases}$$

The diversion rate is given by

$$D_{jk} = \left|\frac{\frac{\partial s_k}{\partial p_j}}{\frac{\partial s_j}{\partial p_j}}\right| = \left|\frac{\frac{\partial s_k}{\partial \delta_j}\frac{\partial \delta_j}{\partial p_j}}{\frac{\partial s_j}{\partial \delta_j}\frac{\partial \delta_j}{\partial p_j}}\right| = \begin{cases} \frac{\rho s_{k|g}+(1-\rho)s_k}{1-\rho s_{j|g}-(1-\rho)s_j}, & k,j \text{ are in the same nest} \\ \frac{(1-\rho)s_k}{1-\rho s_{j|g}-(1-\rho)s_j}, & k,j \text{ are in different nests} \end{cases}$$

Welfare Change In this case, the welfare change in dollars is

$$\Delta CW = \frac{1}{\alpha}\left[\log\left(\sum_{g\in G}\left(\sum_{j\in g}\exp\left(\frac{\delta_j'}{1-\rho}\right)\right)^{1-\rho}\right) - \log\left(\sum_{g\in G}\left(\sum_{j\in g}\exp\left(\frac{\delta_j}{1-\rho}\right)\right)^{1-\rho}\right)\right].$$

2.7 [Berry, Levinsohn, and Pakes(1995)]: Estimation using random-coefficients logit model

Now we consider the case of the random-coefficients logit model.

$$u_{ijt} = x_{jt}\beta_{it} - \alpha_{it}p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

where $(\alpha_{it}, \beta_{it})$ are allowed to vary across i.

We follow [Berry, Levinsohn, and Pakes(1995)] in assuming that each ϵ_{ijt} is an i.i.d. draw from a standard type-1 extreme value (Gumbel) distribution. The most common assumption is that these random variables are jointly normally distributed.

$$(\alpha_{it}, \beta_{it})' \sim N\left((\bar{\alpha}, \bar{\beta})', \Sigma\right)$$

where $\bar{\alpha}$ and $\bar{\beta}$ are the average of α_{it} and β_{it} that represent <u>all</u> consumers' tastes. In this case, define $\delta_{jt} := x_{jt}\bar{\beta} - \bar{\alpha}p_{jt} + \xi_{jt}$. The parameters to be estimated are

$$\theta := (\bar{\alpha}, \bar{\beta}, \Sigma)$$

The choice probabilities take the MNL form:

$$\begin{aligned} s_{jt} &= \int \frac{\exp(x_{jt}\beta_{it} - \alpha_{it}p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J_t} \exp(x_{kt}\beta_{it} - \alpha_{it}p_{kt} + \xi_{kt})} dF(\alpha_{it}, \beta_{it}; t) \\ &= \int \frac{\exp(\delta_{jt} + x_{jt}(\beta_{it} - \bar{\beta}) - (\alpha_{it} - \bar{\alpha})p_{jt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + x_{kt}(\beta_{it} - \bar{\beta}) - (\alpha_{it} - \bar{\alpha})p_{kt})} dF(\alpha_{it}, \beta_{it}; t) \\ &:= \hat{s}_{jt}^{RC}\left(\delta_{1t}, ..., \delta_{J_t t}; \Sigma\right) \end{aligned}$$

Note that the distributions of $\alpha_{it} - \bar{\alpha}$ and $\beta_{it} - \bar{\beta}$ are normal distribution with mean 0 and covariance matrix Σ. So, it can be written as a function that only depends on Σ.

In this case, the inversion step described before will not work, because the J_t equations matching observed to predicted shares have more than J_t unknowns.

[Berry, Levinsohn, and Pakes(1995)] proposed a generalized method of moments (GMM) estimation approach that can be sketched as follows:

1. Take a trial value of the parameters θ;
2. Then, for each market t, the inversion step described before can be used. We "invert" the demand model at the observed market shares $\tilde{s}_{jt}$ to find $\hat{\delta}_{jt}(\Sigma) := \delta_{jt}(\tilde{s}_{1t}, ..., \tilde{s}_{J_t t}; \Sigma)$ as (2.5). The corresponding unobserved characteristics are
$$\xi_{jt}(\theta) := \hat{\delta}_{jt}(\Sigma) - x_{jt}\bar{\beta} + \bar{\alpha}p_{jt}$$
3. Evaluate the trial value θ using a GMM criterion function based on moment conditions (Sample Moment)

of the form

$$\min_{\theta} \frac{1}{N} \sum_{\forall j,t} \xi_{jt}(\theta) z_{jt}$$

where $N = \sum_{t=1}^{T} J_t$ and $z_{jt} \supset x_{jt}$ is a vector of appropriate instrumental variables;

4. Repeat steps 1 and 2 until convergence.

2.8 [Nevo(2001)]: Measuring Market Power

In this paper, he considers different models of supply conduct. Demand parameters are estimated and used to compute the price-cost margins (PCM) implied by different models of conduct.

2.8.1 Supply

Suppose there are F firms. Among all products $\mathcal{J} = \{1, ..., J\}$, the subset of products produced by a firm f is denoted by $\mathcal{F}_f$. The profits of the firm f are

$$\Pi_f = \sum_{j \in \mathcal{F}_f} (p - mc_j) M s_j(p) - C_f$$

where mc_j is the marginal cost, M is the size of the market, $s_j(p)$ is the market share of brand j at price p, and C_f is the fixed cost of production.

Assuming the existence of pure-strategy Bertrand-Nash equilibrium in prices, and that the prices that support it are strictly positive. The price p_j of product j produced by firm f must satisfy the FOC,

$$s_j(p) + \sum_{r \in \mathcal{F}_f} (p_r - mc_r) \frac{\partial s_r(p)}{\partial p_j} = 0 \tag{2.6}$$

By defining $S_{jr} := -\frac{\partial s_r}{\partial p_j}$ and a $J \times J$ matrix Ω such that $\Omega_{jr} := \Omega^*_{jr} S_{jr}$ where

$$\Omega^*_{jr} = \begin{cases} 1, & \text{if } \exists f : \{r, j\} \subset \mathcal{F}_f, \\ 0, & \text{otherwise.} \end{cases}$$

In vector notation, the set of equations of first-order conditions (2.6) can be written as

$$s(p) - \Omega(p - mc) = 0$$

$$\Leftrightarrow \underbrace{p - mc}_{\text{PCM}} = \Omega^{-1} s(p)$$

where $s(\cdot), p$, and mc are $J \times 1$ vectors.

In this paper, there are three hypothetical industry conduct models and the PCM in all of them can be estimated by defining $\mathcal{F}_f$ and Ω^*.

1. Single-product firms: price of each brand is set by a profit-maximizing strategy for that brand.

   ∘: In this case, the PCM arise only from *product differentiation*.

2. Current structure: multi-product firms set prices of all their products jointly.

   ○: In this case, the PCM arise from *product differentiation* and *portfolio effect*.

3. Monopoly / Price collusion: all prices of products are set jointly.

   ○: In this case, the PCM arise from *product differentiation*, *portfolio effect*, and *price collusion*.

2.8.2 Demand

Suppose we observe $t = 1, ..., T$ markets, each with $i = 1, ..., I_t$ consumers. (A market is defined as a city-quarter combination). The conditional indirect utility of consumer i from product j at market t is

$$\begin{gathered} u_{ijt} = x_j \beta_i^* - \alpha_i^* p_{jt} + \xi_j + \Delta\xi_{jt} + \epsilon_{ijt}, \\ \forall i = 1, ..., I_t, j = 1, ..., J_t, t = 1, ..., T, \end{gathered} \tag{2.7}$$

where x_j is a K-dimensional (row) vector of observable product characteristics, p_{jt} is the price of product j in market t, ξ_j is the national mean valuation of the unobserved (by the econometricians) product characteristics, $\Delta\xi_{jt}$ is a city-quarter specific deviation from this mean, and ϵ_{ijt} is a mean-zero stochastic term. Finally, (α_i^*, β_i^*) are $K + 1$ individual-specific *coefficients*.

The distribution of consumers' taste parameters for the characteristics is modeled as multivariate normal (conditional on demographics) with a mean that is a function of demographic variables and parameters to be estimated, and a variance-covariance matrix to be estimated. Let

$$\begin{pmatrix} \alpha_i^* \\ \beta_i^* \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i, \;\; v_i \sim \mathcal{N}(0, I_{K+1}) \tag{2.8}$$

where K is the dimension of the observed characteristics vector, D_i is a $d \times 1$ vector of demographic variables, Π is a $(K + 1) \times d$ matrix of *coefficients* that measure how the taste characteristics vary with demographics, and Σ is a scaling matrix.

The indirect utility from this outside option is

$$u_{i0t} = \xi_0 + \pi_0 D_i + \sigma_0 v_{i0} + \epsilon_{i0t}$$

where ξ_0 is normalized to zero.

Let $\theta = (\theta_1, \theta_2)$ be a vector containing all parameters of the model. The vector $\theta_1 = (\alpha, \beta)$ contains the linear parameters and the vector $\theta_2 = (\text{vec}(\Pi), \text{vec}(\Sigma))$ the nonlinear parameters.

Combining (2.7) and (2.8), we obtain

$$\begin{aligned} u_{ijt} =& \delta_{jt}\left(x_j, p_{jt}, \xi_j, \Delta\xi_{jt}; \theta_1\right) + \mu_{ijt}\left(x_j, p_{jt}, v_i, D_i; \theta_2\right) + \epsilon_{ijt}, \\ \text{where } \delta_{jt} &:= x_j\beta - \alpha p_{jt} + \xi_j + \Delta\xi_{jt}, \\ \mu_{ijt} &:= \begin{bmatrix} p_{jt} \\ x_j \end{bmatrix} \cdot (\Pi D_i + \Sigma v_i) \end{aligned} \tag{2.9}$$

where $D_i, v_i, \epsilon_i t$ are assumed to follow some distributions.

Given $x := (x_1, ..., x_{J_t})^T \in \mathbb{R}^{J_t \times K}, p_{.t} := (p_1, ..., p_{J_t})^T \in \mathbb{R}^{J_t}$, and $\delta_{.t} := (\delta_1, ..., \delta_{J_t})^T \in \mathbb{R}^{J_t}$, the set of (D_i, v_i, ϵ_i) is

$$A_{jt}\left(x, p_{.t}, \delta_{.t}; \theta_2\right) := \{(D_i, v_i, \epsilon_{it}) \mid u_{ijt} \geq u_{ilt}, \forall l = 0, 1, ..., J_t\}$$

Then, the market share of the j^{th} product is give by

$$s_{jt}\left(x, p_{.t}, \delta_{.t}; \theta_2\right) = \int_{A_{jt}} dP^*(D, v, \epsilon) \tag{2.10}$$

where $P^*(\cdot)$ denotes population distribution functions.

2.8.3 Estimation

Let $Z = [z_1, ..., z_M]$ be a set of IV such that $\mathbb{E}[Z^T\omega(\theta^*)] = 0$, where ω is an *error* term that is a function of the model parameters and θ^* denotes the true values of model parameters. The GMM estimate is given by

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \omega(\theta)^T Z A^{-1} Z^T \omega(\theta) \tag{2.11}$$

where A is a constant estimate of $\mathbb{E}[Z^T\omega\omega^T Z]$.

Following [Berry(1994)], the error term ω is defined as the unobserved characteristics, $\xi_j + \Delta\xi_{jt}$. The error term can be obtained by the inversion of the equation system:

$$s_{jt}\left(x, p_{.t}, \delta_{.t}; \theta_2\right) = S_{.t} \tag{2.12}$$

where $S_{.t}$ are the observed market shares. By its inverse function, we can compute δ_{jt} as a function of $S_{.t}$:

$$\delta_{jt}(x, p_{.t}, S_{.t}; \theta_2)$$

Then, the error term is defined as

$$\omega_{jt} = \delta_{jt}(x, p_{.t}, S_{.t}; \theta_2) - (x_j\beta + \alpha p_{jt})$$

And the A in (2.11) is computed by the two-step procedure:

1. Set the weight matrix to $Z^T Z$ and compute the initial estimate of the parameters, denoted by $\theta^{(1)}$:

$$\theta^{(1)} = \operatorname*{argmin}_{\theta} \omega(\theta)^T \omega(\theta)$$

2. Use the initial estimate to re-compute the weight matrix, i.e., $A = \frac{1}{n}\sum_{i=1}^{n} Z^T\omega(\theta^{(1)})\omega(\theta^{(1)})^T Z$, where

n is the number of observations (i.e., consumers).

3. We can repeat the step 1 to update the estimate until convergence. (In this paper, it does not because the estimated parameters changed only slightly beyond the second iteration).

With the weight matrix A, we can solve the (2.11). In the full random coefficients model, both the computation of the market shares, and the inversion in order to get $\delta_{jt}(\cdot)$, have to be done numerically. Then, (2.11) is solved by nonlinear search...

Brand-Specific Dummy Variables This paper includes brand-specific dummy variables.

Instruments The prices of the brand in other cities are valid IV's: the prices of brand j in two cities are correlated due to the common marginal cost, but are uncorrelated with market-specific valuation due to the independence assumption.

Chapter 3 Estimation with Extra Data

3.1 [Petrin(2002)]: Using demographic data of consumers

[Berry, Levinsohn, and Pakes(1995)] allows substitution patterns to reflect consumer-level heterogeneity in tastes for observed product characteristics. Modeling this heterogeneity is important when estimating demand elasticities, although the estimates tend to be imprecise when constrained to market-level data. [Goldberg(1995)] and [Berry, Levinsohn, and Pakes(2004)] utilize consumer-level data to improve these estimates.

This paper offer an alternative approach when consumer-level data are not available: relate the average demographics of consumers to the characteristics of the products they purchase. The extra information plays the same role as consumer-level data, allowing estimated substitution patterns and (thus) welfare to directly reflect demographic-driven differences in tastes for observed characteristics.

3.1.1 Demand

The conditional indirect utility given observed and unobserved product j and consumer i characteristics and model parameters $\vec{\theta}$ is given as

$$u_{ij}(\vec{\theta}) = \delta_j(\vec{\theta}) + \mu_{ij}(\vec{\theta}) + \epsilon_{ij}.$$

The first component, $\delta_j(\vec{\theta})$, is a product-specific term common to all consumers. The $\mu_{ij}(\vec{\theta})$ term captures heterogeneity in consumer tastes for observed product characteristics.

The utility component common to all consumers, δ_j is usually given as

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

where X_j is the vector of product characteristics.

The utility in this paper is specified as

$$u_{ij} = \alpha_i \ln(y_i - p_j) + X_j\beta + \sum_k \underbrace{\gamma_k v_{ik}}_{:=\gamma_{ik}} x_{jk} + \xi_j + \epsilon_{ij},$$

where $v_i = (v_{i1}, ..., v_{iK})$ is the consumer i's K idiosyncratic tastes for the K observed characteristics, and γ_k is a parameter measuring the heterogeneity in tastes for the observed characteristics in the population. We also use $\gamma_{ik} := \gamma_k v_{ik}$ to measure consumer i's personal taste for characteristic k.

The marginal utility of income to vary according to income groups:

$$\alpha_i = \begin{cases} \alpha_0, & \text{if } y_i < \bar{y}_1 \\ \alpha_1, & \text{if } \bar{y}_1 \leq y_i \leq \bar{y}_2 \end{cases}$$

where $\bar{y}_1$ and $\bar{y}_2$ divide the U.S. population into three equally sized groups ordered by income.

Bibliography

[Angrist, Graddy, and Imbens(2000)] Angrist, J. D., Graddy, K., and Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies*, 67(3):499–527.

[Berry, Levinsohn, and Pakes(1995)] Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

[Berry, Levinsohn, and Pakes(2004)] Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy*, 112(1):68–105.

[Berry(1994)] Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

[Berry and Haile(2021)] Berry, S. T. and Haile, P. A. (2021). Foundations of demand estimation. In *Handbook of industrial organization*, volume 4, pages 1–62. Elsevier.

[Goldberg(1995)] Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the us automobile industry. *Econometrica: Journal of the Econometric Society*, pages 891–951.

[McFadden(1972)] McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior.

[McFadden(1981)] McFadden, D. (1981). Econometric models of probabilistic choice, structural analysis of discrete data with econometric applications. *Cambridge: MIT press*.

[Nevo(2001)] Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

[Petrin(2002)] Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, 110(4):705–729.

[Working(1927)] Working, E. J. (1927). What do statistical "demand curves" show? *The Quarterly Journal of Economics*, 41(2):212–235.