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

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

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

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

2.1 Random Utility Models

Let $j = 1, ..., J_i$ 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 2.1 (Random Utility Model)**
>
> Given the choice set (J_i, χ_i), each consumer's utility vector $(u_{ij})_{j=0,1,...,J_i}$ 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 $(q_{i1}, ..., q_{iJ_i})$, where

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

The consumer-specific choice probabilities are then given by

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

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

2.2 The Canonical Model

> **Definition 2.2 (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}.

2.3 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.

2.4 [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.

2.4.1 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

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

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 2.3

[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:

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

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

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

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 $\ln(s_1 + \epsilon) - \ln(s_2 + \epsilon) \neq \ln(s_1) - \ln(s_2)$.

2.4.2 Supply Side Recover

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

$$D^j\left(\vec{x}, \vec{p}, \vec{\xi}\right)$$

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:

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

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

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

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

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

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

where $\vec{D} := (D_1, ..., D_J)$ and ΔD is a $J \times J$ matrix where

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

2.5 [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:

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

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.

2.6 [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.

2.6.1 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

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

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*.

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(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.

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

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