Demand Estimation 3
PhD Industrial Organization

Nicholas Vreugdenhil

Plan

1. Identification: what if we had micro-data?
2. Identification: Σ
3. Estimation algorithm

Plan

1. **Identification: what if we had micro-data?**

2. Identification: Σ

3. Estimation algorithm

Review: setup of the problem

- What are the parameters we need to estimate?
- Linear parameters:
  - Parameters from the mean utility equation: (α_0, β_0)
- Nonlinear parameters
  - Γ: coefficients on (observed) demographics
  - Σ: idiosyncratic "taste for characteristics"
- So, full parameter vector to estimate: $\theta = (\alpha_0, \beta_0, \Gamma, \Sigma)$.

Review: Identification

- What variation in the data can identify the parameters?
  - Precise econometric definition of identification: see haile.pdf on Canvas, or 'The Identification Zoo' (Lewbel)

- **Thought experiment**: what if we:
  - 1. have micro-data on individual consumers
  - 2. observe a single market
  - 3. switch off $\Sigma = 0$ (i.e. ignore any idiosyncratic "taste for characteristics", implies heterogeneity is only driven by observed demographics)

- Later, we will build on this intuition to discuss what to do if we had more aggregated market-level data with random taste shocks etc...

Review: Identification using individual-level data

- (Conditional indirect) utility from product j (dropping t subscript and incorporating price p_j as a 'characteristic' in x_j to simplify exposition):

$$u_{ij} = \underbrace{x_j\beta_0 + \xi_j}_{\delta_j} + \sum_{k,l} \beta_d^{(l,k)} D_{il} x_{jk} + \epsilon_{ij}$$

- Instead, use a **two-step procedure**:

- 1. Include a product-specific intercept to capture $\delta = x_j\beta_0 + \xi_j$ (i.e. estimate $\tilde{\theta} = (\delta_1, ..., \delta_J, \Gamma)$ using maximum likelihood)

- 2. Estimate β_0 by 'projecting' estimated δ's on the x's.

Identification using individual-level data: step 1

- Identifying Γ:
- Again, looking at FOC from maximum likelihood, can show that estimates of Γ:
- Equate observed to predicted covariance between demographic variables of consumers who choose product j and the characteristics of the product j.
- Asymptotically, Γ solves $L(K+1)$ equations given by:

$$E_{Population}[x^k D^l] = E_{Model}[x^k D^l; \Gamma]$$

Identification using individual-level data: what if $\Sigma \neq 0$?

- What if we had a more complicated model where $\Sigma \neq 0$?
  - i.e. still one market, but also include unobserved heterogeneity for idiosyncratic "tastes in characteristics"

- Now consider the first step of the two-step procedure from before:
  - For (δ, Γ), first order conditions still hold.
  - But, Σ and Γ are **not** separately identified in this particular thought experiment.
    - For Σ, we have additional moment conditions related to the covariance. These look very similar to the moment conditions for Γ.
    - So, it's not clear if these covariance moments are identifying the Σ parameters or the Γ parameters, since ν is unobserved.

- However, we can use variation in the second-stage moments from before (I now explain this in detail over the next few slides...)

Plan

1. Identification: what if we had micro-data?
2. **Identification:** Σ
3. Estimation algorithm

Identifying Σ

- **Thought experiment**: what if we:
  - 1. have only market-level data
  - 2. observe a single market
  - 3. switch off $\Gamma = 0$ (this is just for exposition)
- Then (conditional indirect) utility is:

$$u_{ij} = \delta_j + \sum_k \beta_\nu^{(k)} \nu_{ik} x_{jk} + \epsilon_{ij}$$

- **Question:** how do we pin down (δ, Σ)?

Identifying Σ

- Can aggregate market share data alone separately identify δ and Σ?

Identifying Σ

- Can aggregate market share data alone separately identify δ and Σ?
  - No...
  - For any given Σ we can choose mean utilities δ that exactly equate predicted shares to observed shares using the 'Berry inversion' from before.
  - No variation left in the data to pin down Σ

Identifying Σ

- What if we **also** had additional moment restrictions from 'step 2' from before: $E[\xi_j|\mathbf{Z}] = 0$?
- We will work with the common assumption that $\mathbf{Z} = \mathbf{x}$ i.e. $\mathbf{Z}$ stacks all the product characteristics
    - In words: means 'unobserved component of mean utility is mean-independent of market structure
    - Usually we would exclude price and advertising from these product characteristics due to endogeneity concerns'
    - Later we will talk about *why* this assumption might be justified. Let's just see what it does for now...

Identifying Σ using $E[\xi_j|\mathbf{Z}] = 0$

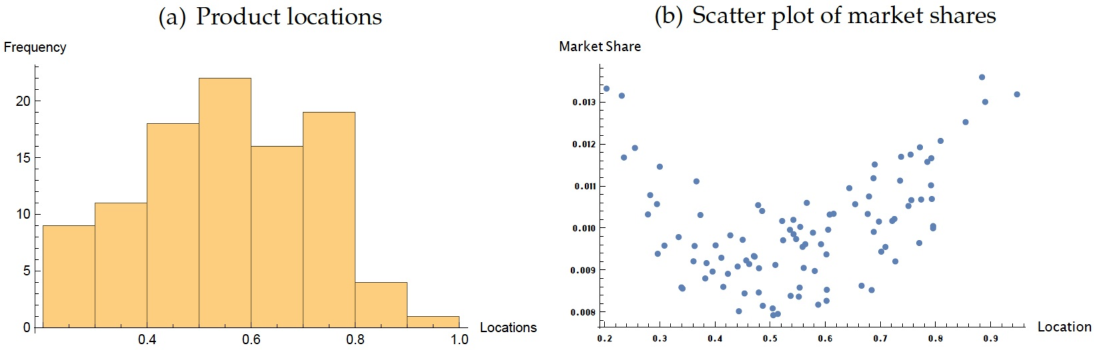
- Consider one-dimensional Hotelling model. Utility to i for product j:

$$u_{ij} = -\theta \cdot d(t_i, x_j) + \xi_j + \epsilon_{ij}$$

- Here:
  - θ: travel cost (this takes the place of our 'unobserved taste for characteristics'). Assume $\theta \geq 0$.
  - d: distance between location $t_i \in [0, 1]$ of consumer i and location $x_j \in [0, 1]$ of product j.
  - ξ_j: mean quality of product j
  - ϵ_{ij}: idiosyncratic taste shocks drawn from type-1 extreme value distribution (i.e. logit draws)
  - draw 100 product locations x_j from a Beta distribution

Identifying Σ using $E[\xi_j|\mathbf{Z}] = 0$

Figure 4.1: Distribution of product locations and market shares



- Panel (a): bunching towards center (just follows from the Beta distribution)

- Panel (b): in crowded parts of product space market shares are relatively smaller. Why?

Identifying Σ using $E[\xi_j|\mathbf{Z}] = 0$

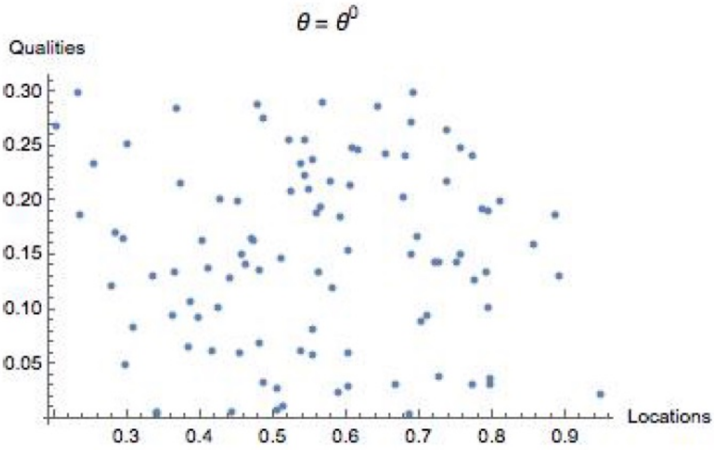
- Two possibilities could rationalize the "data" on market share patterns in panel (b):
  - 1. Travel costs θ are large, so most products compete locally
  - 2. Travel costs $\theta = 0$, but products that are located in the center have *systematically lower qualities* ξ_j
- **This is the intuition behind why market share data alone cannot distinguish between unobserved tastes** (θ) **vs unobserved quality** (ξ_j).

Identifying Σ using $E[\xi_j|\mathbf{Z}] = 0$

- Let's now consider what happens if we also have moments $E(\xi_j|x_j) = 0$.
- On next slides: θ_0 is the 'true' value in the model. The value θ is an alternative 'guess' of θ_0 that may or may not be different from the 'truth'
  - i.e. obtain these by guessing θ and then getting ξ_j that fit observed market shares.
  - The graphs on the next slides plot the quality $\xi_j(\theta)$ (implied by the model) vs the product location x_j.
- As we will see, the $E(\xi_j|x_j) = 0$ moment rules out the second explanation from before.

Identifying Σ using $E[\xi_j|\mathbf{Z}] = 0$

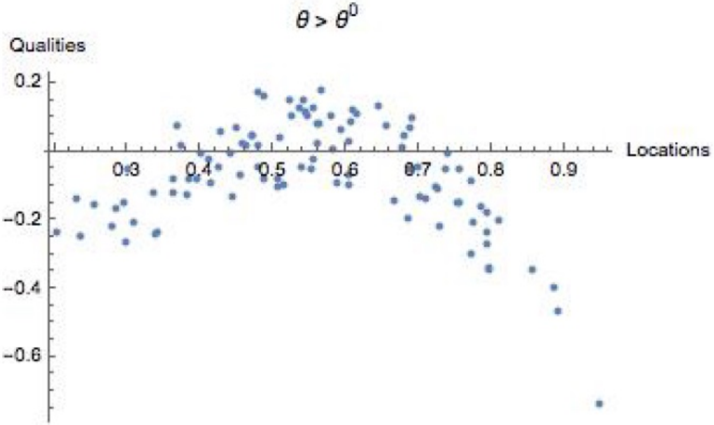
(a) $\theta = \theta^0$



- If $\theta = \theta_0$, (implied) quality is uncorrelated with location.

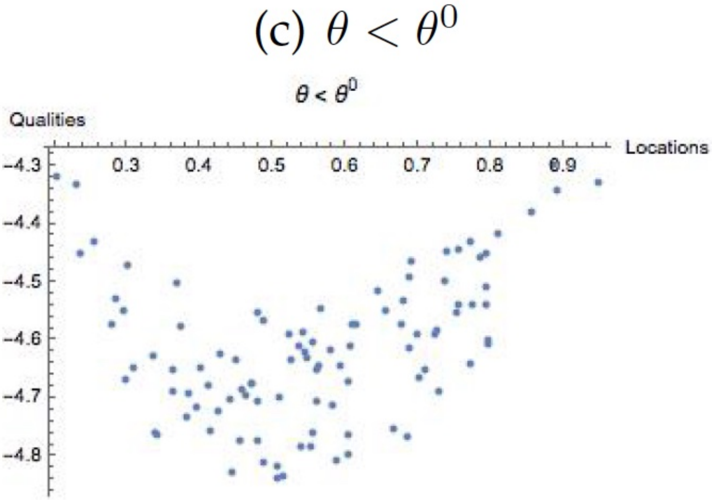
Identifying Σ using $E[\zeta_j|\mathbf{Z}] = 0$

(b) $\theta > \theta^0$



- If $\theta > \theta_0$, data exhibit correlation.

Identifying Σ using $E[\zeta_j|\mathbf{Z}] = 0$



- If $\theta < \theta_0$, data exhibit correlation.

Identifying Σ using $E[\xi_j|\mathbf{Z}] = 0$

- Main takeaways from this exercise:
- 1. **Two roles** for IVs in the model $E[\xi_j|\mathbf{Z}] = 0$: dealing with price endogeneity and identifying non-linear parameters Σ
    - This is under-appreciated!
    - If we had micro-data we would know more moments of the distribution of choice probabilities for each product which would also be helpful for identification.
- 2. Parameters of the model could potentially be identified from aggregate data on a single market.
    - Often, we will also have cross-market data.

Plan

1. Identification: what if we had micro-data?
2. Identification: Σ
3. **Estimation algorithm**

Estimation algorithm

- I will focus on the 'nested fixed point' algorithm used in the original BLP paper.
  - We will allow for both within- and across-market variation.
  - Continue to assume we have the moment restrictions: $E[\xi_{jt}|\mathbf{Z}_t] = 0$.
  - We will talk about the choice of IVs later.
- We will then discuss alternative approaches.

Estimation algorithm: overview

- Preliminary:
  - Get (and fix) R random draws from $F_\nu(\nu)$ (usually a standard normal) and $\hat{F}_D$ (e.g. an empirical distribution of demographics from Census data).
  - Also convert quantities to market shares by dividing by an assumed market size.
- Step 1: For a guess of Γ and Σ, and a vector of mean utilities $\boldsymbol{\delta}_t$, compute model-predicted market shares.
- Step 2: For a guess of Γ and Σ do an **inversion**: find $\boldsymbol{\delta}_t$ where the model-predicted market shares match the empirical market shares s_t.
  - This step will repeatedly call the function from Step 1.
- Step 3: Use the computed $\boldsymbol{\delta}_t$ from Step 2 to compute $\xi_{jt} = \delta_{jt}(\Gamma, \Sigma) - x_{jt}\beta_0 - \alpha_0 p_{jt}$.
  - Interact with IVs to get the GMM objective function.
  - Search over all parameters θ to minimize objective function using non-linear optimization.

Estimation algorithm: simple example

- What if we have a logit model? $(\Gamma = 0, \Sigma = 0)$

- Step 1:

$$s_{jt} = \frac{\exp\{\delta_{jt}\}}{\sum_{k=0}^{J} \exp\{\delta_{kt}\}}$$

- Step 2:

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} - \underbrace{\delta_{0t}}_{=0} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$$

- Step 3: Estimate above equation with e.g. 2SLS.
  - In other words, you can collapse the estimation down to a single step for the logit example.
  - To make sure you understand: think about how you would construct the dependent variable 'data' in this single step.

Estimation algorithm: preliminaries

- Get (and fix) R random draws from $F_\nu(\nu)$ (usually a standard normal) and $\hat{F}_D$ (e.g. an empirical distribution of demographics from Census data).
    - Denote these draws $\hat{F} = \{\hat{\nu}_{it}, \hat{D}_{it}\}_{i=1}^{R}$
- Also convert quantities to market shares by dividing by an assumed market size (I_t):
    - i.e. $s_{jt} = q_{jt} / I_t$.

Estimation algorithm: step 1

- For a guess of Γ and Σ, and a vector of mean utilities δ_t, compute model-predicted market shares.

$$\tilde{\sigma}(\delta_t; \Gamma, \Sigma, \mathbf{x}_t, \mathbf{p}_t, \hat{F}) = \frac{1}{R} \sum_{i=1}^{R} \frac{\exp\{\delta_{jt} + (x_{jt}, p_{jt}) \cdot (\Gamma D_{it} + \Sigma \nu_{it})\}}{1 + \sum_{k=1}^{J} \exp\{\delta_{kt} + (x_{kt}, p_{kt}) \cdot (\Gamma D_{it} + \Sigma \nu_{it})\}}$$

- Can compute this using the simulated draws from the first step.

- Could also use e.g. quadrature methods (since we are computing an integral)

- Many tricks used to speed up this step in the literature (e.g. vectorization). See Conlon and Gortmaker (2020).

Estimation algorithm: step 2

- For a guess of Γ and Σ do an **inversion**: find δ_t where the model-predicted market shares match the empirical market shares s_t:
- (i) Find a starting guess of δ_t
- (ii) Update in the following way (Berry (1994) shows that this is a contraction mapping):

$$\delta_t^{r+1} = \delta_t^r + \ln(\mathbf{s}_t) - \ln \tilde{\sigma}(\delta_t^r; \Gamma, \Sigma, \mathbf{x}_t, \mathbf{p}_t, \hat{F})$$

- (iii) Continue iterating until $||\delta_t^{r+1} - \delta_t^r|| < \tau$ where τ is very small.
  - Note that τ needs to be <u>really</u> small (e.g. 10^{-12}).
  - Knittel and Metaxoglou (2014) document that the original BLP paper actually gets the estimates wrong because the tolerances are too loose.

Estimation algorithm: step 3

- Denote the mean utilities from step 2: $\delta_{jt}(\Gamma, \Sigma)$
- Compute $\xi_{jt}(\theta) = \delta_{jt}(\Gamma, \Sigma) - x_{jt}\beta_0 - \alpha_0 p_{jt}$
  - Above equation is why we called Γ, Σ 'nonlinear' variables, and β_0, α_0 the 'linear variables'
- Interact with the instrumental variables to get the GMM objective function (denoting W as the GMM weight matrix):

$$\xi(\theta)' ZWZ' \xi(\theta)$$

- Solve for the parameters using nonlinear optimization.

$$\hat{\theta} = arg \min_{\theta} \xi(\theta)' ZWZ' \xi(\theta)$$

- Note: since this is just a GMM problem, can also get standard errors using standard GMM methods