

Aggregate Data

Chris Conlon

Saturday 4th October, 2025

Grad IO

- ▶ Now we want to have both **price endogeneity** and **flexible substitution** in the same model.
- ▶ We are ultimately going with the random coefficients logit model, but we will start with the logit and nested logit.
- ▶ We will explore a technique that works with **aggregate data**.

Multinomial: Aggregation Property

Each individual's choice $y_{ij} \in \{0, 1\}$ and $\sum_{j \in \mathcal{J}} d_{ij} = 1$.

Choices follow a Multinomial distribution with $m = 1$:

$$(d_{i1}, \dots, d_{iJ}, d_{i0}) \sim \text{Mult}(1, s_{i1}, \dots, s_{iJ}, s_{i0})$$

If each individual faces the same $s_{ij} = s_j$ the the sum of Multinomials is itself Multinomial:

$$(q_1^*, \dots, q_J^*, q_0^*) \sim \text{Mult}(M, s_1, \dots, s_J, s_0)$$

where $q_j^* = \sum_{i=1}^M d_{ij}$ is a **sufficient statistic**.

Multinomial: Aggregation Property (Likelihood)

We can write the likelihood as $L((y_{i1}, \dots, y_{iJ}, y_{i0}) \mid \mathbf{x}_i, \theta)$ where \mathbf{x}_i is a J vector that includes all relevant product characteristics interacted with all relevant individual characteristics.

$$\begin{aligned} &= \binom{M}{q_{i1}, \dots, q_{iJ}, q_{i0}} \prod_{i=1}^M s_{i1}(\mathbf{x}_i, \theta)^{d_{i1}} \dots s_{iJ}(\mathbf{x}_i, \theta)^{d_{iJ}} s_{i0}(\mathbf{x}_i, \theta)^{d_{i0}} \\ \rightarrow \ell(\mathbf{x}_i, \theta) &= \sum_{i=1}^M \sum_{j \in \mathcal{J}} d_{ij} \log s_{ij}(\mathbf{x}_i, \theta) + \log C(\mathbf{q}) \end{aligned}$$

If all individuals face the same (\mathbf{x}_i) and \mathcal{J} they will have the same $s_{ij}(\mathbf{x}_i, \theta)$ and we can aggregate outcomes into **sufficient statistics**.

$$\rightarrow \ell(\theta) = \sum_{j \in \mathcal{J}} q_j^* \log s_j(\theta)$$

Aggregation is probably the most important property of discrete choice:

- ▶ Instead of individual data, or a single group we might have multiple groups: if prices only change once per week, we can aggregate all of the week's sales into one "observation".
- ▶ Likewise if we only observe that an individual is within one of five income buckets – there is no loss from aggregating our data into these five buckets.
- ▶ All of this depends on the precise form of $s_{ij}(\mathbf{x}_i, \theta)$. When it doesn't change across observations: we can aggregate.
- ▶ Notice I didn't need anything to follow a logit/probit.

$$s_{ij}(\mathbf{x}_i, \theta) = \int \frac{\exp[x_{ij}\beta_\iota]}{1 + \sum_k \exp[x_{ik}\beta_\iota]} f(\beta_\iota|\theta) \partial\beta_\iota = \sum_{\iota=1}^S w_\iota \frac{\exp[x_{ij}\beta_\iota]}{1 + \sum_k \exp[x_{ik}\beta_\iota]}$$

- ▶ Notice that while i subscripts “individuals” with different characteristics \mathbf{x}_i
- ▶ ι is the dummy index of integration/summation.
 - Even though we sometimes call these “simulated individuals”
 - Everyone with the same \mathbf{x}_i still has the same $s_{ij}(\mathbf{x}_i, \theta)$
- ▶ Most papers will abuse notation and i will serve double duty!

Now suppose we have aggregate data: (q_1, \dots, q_J, q_0) where $M = \sum_{j \in \mathcal{J}} q_j$.

- ▶ If M gets large enough then $(\frac{q_1}{M}, \dots, \frac{q_J}{M}, \frac{q_0}{M}) \rightarrow (\mathfrak{s}_1, \dots, \mathfrak{s}_J, \mathfrak{s}_0)$
 - Idea: Observe $(\mathfrak{s}_1(\mathbf{x}_i), \dots, \mathfrak{s}_J(\mathbf{x}_i), \mathfrak{s}_0(\mathbf{x}_i))$ without sampling variance.
 - Challenges: We probably don't really observe q_0 and hence M .
- ▶ Idea: Equate observed market shares to the conditional choice probabilities $(s_1(\mathbf{x}_i, \theta), \dots, s_J(\mathbf{x}_i, \theta), s_0(\mathbf{x}_i, \theta))$.
- ▶ Choose θ that minimizes distance: MLE? MSM? Least Squares? etc.

Inversion: IIA Logit

Add unobservable error for each s_{jt} labeled ξ_{jt} .

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \varepsilon_{ijt}, \quad s_{jt} = \frac{\exp[x_{jt}\beta - \alpha p_{jt} + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta - \alpha p_{kt} + \xi_{kt}]}$$

- ▶ The idea is that ξ_{jt} is observed to the firm when prices are set, but not to us the econometricians.
- ▶ Potentially correlated with price $\text{Corr}(\xi_{jt}, p_{jt}) \neq 0$
- ▶ But not characteristics $E[\xi_{jt}|x_{jt}] = 0$.
 - This allows for products j to better than some other product in a way that is not fully explained by differences in x_j and x_k .
 - Something about a BMW makes it better than a Peugeot but is not fully captured by characteristics that leads higher sales and/or higher prices.
 - Consumers agree on its value (vertical component).

Taking logs:

$$\ln s_{0t} = -\log \left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}] \right)$$

$$\ln s_{jt} = [x_{jt}\beta - \alpha p_{jt} + \xi_{jt}] - \log \left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}] \right)$$

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{\text{Data!}} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

Exploit the fact that:

1. $\ln s_{jt} - \ln s_{0t} = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t}$ (with no sampling error)
2. We have one ξ_{jt} for every share s_{jt} (one to one mapping)

IV Logit Estimation (Berry 1994)

1. Transform the data: $\ln s_{jt} - \ln s_{0t}$.
2. IV Regression of: $\ln s_{jt} - \ln s_{0t}$ on $x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ with IV z_{jt} .

Was it magic?

- ▶ No. It was just a nonlinear change of variables from $s_{jt} \rightarrow \xi_{jt}$.
- ▶ Our (conditional) moment condition is just that $E[\xi_{jt}|x_{jt}, z_{jt}] = 0$.
- ▶ We moved from the space of shares and MLE for the logit to the space of utilities and an IV model.
 - We are losing some efficiency – but now we are able to estimate under weaker conditions.
 - But we need **aggregate data** and shares without sampling variance.

Did we need to do change of variables? Imagine we work with:

$$s_{jt} = \frac{\exp[x_{jt}\beta - \alpha p_{jt}]}{1 + \sum_k \exp[x_{kt}\beta - \alpha p_{kt}]}$$
$$\eta_{jt} \equiv (s_{jt}(\theta) - \mathfrak{s}_{jt})$$

- ▶ Each share depends on all prices (p_{1t}, \dots, p_{Jt}) and characteristics \mathbf{x}_t .
- ▶ Harder to come up with IV here.

Inversion: Nested Logit (Berry 1994 / Cardell 1991)

This takes a bit more algebra but not much

$$\underbrace{\ln s_{jt} - \ln s_{0t} - \rho \log(s_{j|gt})}_{\text{data!}} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

$$\ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t} = x_{jt}\beta - \alpha p_{jt} + \rho \log(\mathfrak{s}_{j|gt}) + \xi_{jt}$$

- ▶ Same as logit plus an extra term $\log(s_{j|g})$ the **within group share**.
 - We now have a second endogenous regressor.
 - If you don't see it – realize we are regressing Y on a function of Y . This should always make you nervous.
- ▶ If you forget to instrument for ρ you will get $\rho \rightarrow 1$ because of **attenuation bias**.
- ▶ A common instrument for ρ is the number of products within the nest. Why?

Think about a **generalized inverse** for $\sigma_j(\mathbf{x}_t, \theta_2) = \mathfrak{s}_{jt}$ so that

$$\sigma_{jt}^{-1}(\mathcal{S}_{.t}, \tilde{\theta}_2) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

- ▶ After some transformation of data (shares $\mathcal{S}_{.t}$) we get **mean utilities** δ_{jt} .
 - We assume $\delta_{jt} = h(x_{jt}, v_{jt}, \theta_1) - \alpha p_{jt} + \xi_{jt}$ follows some parametric form (often linear).
- ▶ Same IV-GMM approach after transformation
- ▶ Examples:
 - Plain Logit: $\sigma_j^{-1}(\mathcal{S}_{.t}) = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t}$
 - Nested Logit: $\sigma_j^{-1}(\mathcal{S}_{.t}, \rho) = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t} + \rho \ln \mathfrak{s}_{j|gt}$

Inversion: BLP (Random Coefficients)

We can't solve for δ_{jt} directly this time.

$$\sigma_j(\boldsymbol{\delta}_t, \tilde{\theta}_2) = \int \frac{\exp[\delta_{jt} + \mu_{ijt}]}{1 + \sum_k \exp[\delta_{kt} + \mu_{ikt}]} f(\boldsymbol{\mu}_{it} | \tilde{\theta}_2)$$

- ▶ This is a $J \times J$ system of equations for each t .
- ▶ It is diagonally dominant (with outside good).
- ▶ There is a unique vector $\boldsymbol{\delta}_t$ that solves it for each market t .
- ▶ If you can work out $\frac{\partial s_{jt}}{\partial \delta_{kt}}$ (easy) you can solve this using Newton's Method.

Contraction: BLP

BLP actually propose an easy solution to find δ_t . Fix $\tilde{\theta}_2$ and solve for δ_t . Think about doing this one market at a time:

$$\delta_t^{(k)}(\tilde{\theta}_2) = \delta_t^{(k-1)}(\tilde{\theta}_2) + [\log(s_j) - \log(s_j(\delta_t^{(k-1)}, \tilde{\theta}_2))]$$

- ▶ They prove (not easy) that this is a **contraction mapping**.
- ▶ If you keep iterating this equation enough $\|\delta_t^{(k)}(\theta) - \delta_t^{(k-1)}(\theta)\| < \epsilon_{tol}$ you can recover the δ 's so that the observed shares and the predicted shares are identical.
- ▶ Practical tip: ϵ_{tol} needs to be as small as possible. ($\approx 10^{-13}$).
- ▶ Practical tip: Contraction isn't as easy as it looks: $s_j(\delta_t^{(k-1)}, \tilde{\theta}_2)$ requires computing the numerical integral each time (either via quadrature or monte carlo).

From the outside, in:

- ▶ Outer loop: search over nonlinear parameters θ to minimize GMM objective:

$$\widehat{\theta}_{BLP} = \arg \min_{\theta_2} (Z' \hat{\xi}(\theta_2)) W (Z' \hat{\xi}(\theta_2))'$$

- ▶ Inner Loop:

- Fix a guess of $\tilde{\theta}_2$.
- Solve for $\delta_t(\mathcal{S}_t, \tilde{\theta}_2)$ which satisfies $\sigma_{jt}(\delta_t, \tilde{\theta}_2) = s_{jt}$.
 - Computing $s_{jt}(\delta_t, \tilde{\theta}_2)$ requires numerical integration (quadrature or monte carlo).
- We can do IV-GMM to recover $\hat{\alpha}(\tilde{\theta}_2), \hat{\beta}(\tilde{\theta}_2), \hat{\xi}(\tilde{\theta}_2)$.

$$\delta_t(\mathcal{S}_t, \tilde{\theta}_2) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

- Use $\hat{\xi}(\theta)$ to construct sample moment conditions $\frac{1}{N} \sum_{j,t} Z'_{jt} \xi_{jt}$

- ▶ When we have found $\hat{\theta}_{BLP}$ we can use this to update $W \rightarrow W(\hat{\theta}_{BLP})$ and do 2-stage GMM.

- ▶ Extensions and Variants
- ▶ Supply Side Restrictions
- ▶ Instruments
- ▶ Implementation Details