Choice Models in Marketing

Capturing heterogeneity

MS Seminar

Random Utility Models

- The indirect utility of alternative j is U_j
- $U_{ijt} = V_{ijt} + \varepsilon_{ijt}$, $V_0 = 0$

$$U_{ijt} = \alpha_{ij} - \beta_1 price_{ijt} + \beta_2 adv_{ijt} + \beta_3 display_{ijt} + \varepsilon_{ijt}$$

- Alternative with largest utility is chosen

$$P(choice_{it} = j) = prob(V_{ijt} + \varepsilon_{ijt} > V_{ikt} + \varepsilon_{ikt}, \forall k \neq j) =$$

$$prob(\varepsilon_{ijt} - \varepsilon_{ikt} > V_{ikt} - V_{ijt}, \forall k \neq j)$$

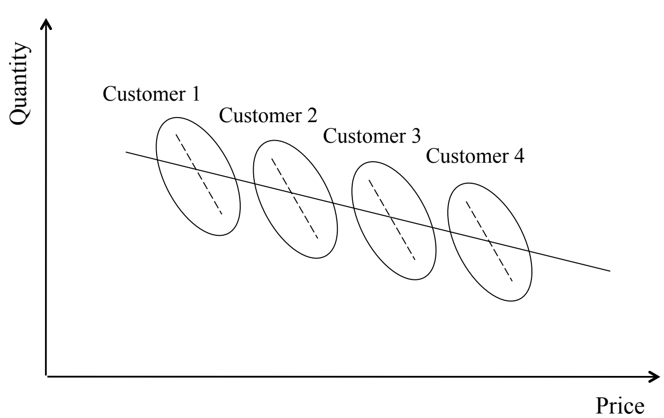
- Utilities are relative - “only differences matter”
- Probit $\varepsilon_{ijt} : N(0,\sigma^2)$ $\boldsymbol{\varepsilon}_{\mathbf{ijt}} : N(0,\Sigma)$
- Logit $\varepsilon_{ijt} :$ Extreme Value

Capturing Heterogeneity

$$U_{ijt} = \alpha_{ij} - \beta_{i1} price_{ijt} + \beta_{i2} adv_{ijt} + \beta_{i3} display_{ijt} + \varepsilon_{ijt}$$

- Consumers differ in their βs
- What are the difficulties?
- What do we need to identify these differences?
- Why is it important to account for heterogeneity?

Capturing Heterogeneity



Capturing Heterogeneity

The effect of price and adv. type on brand attitude

	Price high	Price low
Image appeal	R_{11k}	R_{12k}
Quality appeal	R_{21k}	R_{22k}

$$R_{ijk} = \beta_{0k} + \beta_{rk}X_r(i) + \beta_{ck}X_c(j) + \beta_{rck}X_r(i)X_c(j) + \epsilon_{ijk},$$

	Price high	Price low	Row average	Intercept, β_0	Row, β_r	Column, β_c	Interaction, β_{rc}
				Coefficients ignoring gender			
B. Aggregate results: Total sample (n = 48 per cell)							
Image appeal	3.0	3.0	3.0				
Quality appeal	3.0	3.0	3.0	3.0	.0	.0	.0
C. Unobserved segments:							
Males (n= 24 per cell)				Male coefficients			
Image appeal	3.0	.0	1.5				
Quality appeal	6.0	5.0	5.5				
Column average	4.5	2.5	3.5	3.5	−2.0	1.0	.5
Females (n= 24 per cell):				Female coefficients			
Image appeal	3.0	6.0	4.5				
Quality appeal	.0	1.0	.5				
Column average	1.5	3.5	2.5	2.5	2.0	−1.0	−.5

Source: Hutchinson, Kamakura and Lynch, JCR 2000

Methods to Capture Heterogeneity

- Observed variables (e.g., demographics, loyalty)
- Fixed effects
- Random effects
  - Discrete –
    - Latent Class Models/ Finite Mixture Models
  - Continuous -
    - Hierarchical Bayes
    - Simulated Maximum Likelihood

Methods to Capture Heterogeneity

	Individual	Group
Observed Hetero.	Fixed effects Interaction effect	Interaction with grouping variables
Unobserved Hetero.	Random effect	Latent class

Latent Class

- Finite number of segments
- Each segment has a different set of coefficients
- Distribution of β_i across customers can be approximated by a discrete distribution
  - β_i takes S distinct value for each support
- Advantages?
- Disadvantages?

Latent Class - Estimation

- Use data to figure out:
  - # of segments
  - Coefficients
  - Segment membership
  - Segment size

Latent Class - Estimation

1. Compute the choice probabilities for each purchase occasion for each customer assuming that the customer belongs to each of the S segments
2. Conditional on belonging to segment s, write the customer likelihood function (product of the probability of the chosen brand across all occasions)
3. Compute the weighted sum of likelihood across all segments
4. Compute the product of the likelihoods across customers

Latent Class - Estimation

$$L_{i|s} = \prod_{t=1}^{T_i} \prod_{j=1}^{J} \text{Prob}(Choice_{it|s} = j)^{Y_{ijt}}$$

$$LL = \log\left(\prod_{i=1}^{N} \sum_{s=1}^{S} L_{i|s} \pi_s \right) = \sum_{i=1}^{N} \log\left(\sum_{s=1}^{S} L_{i|s} \pi_s \right)$$

$$0 \le \pi_s \le 1; \quad \sum_{s=1}^{S} \pi_s = 1$$

Latent Class - Estimation

- Determining the # of segments

  BIC=-2LL+(# of param.)*log(# of obs.)

  AIC=2* (# of param.)-2LL
- Computing elasticities
- Concomitant variables - can make π_s function of individual characteristics
- Posterior segment membership probabilities

prior posterior

$$\pi_i = \frac{\pi_s L_{is}}{\sum_{q=1}^{S} \pi_s L_{iq}}$$

Continuous Heterogeneity Distribution

- Distribution of β_i across customers follows a multivariate distribution (e.g., MVN)
- Estimate the mean of β and the covariance matrix associated with it
- No closed-form expression
  - Simulated Maximum Likelihood – SML
  - Bayesian estimation

Mixed Logit

$$U_{ijt} = \boldsymbol{\beta}'_{\mathbf{i}}\mathbf{X}_{\mathbf{ijt}} + \varepsilon_{ijt}$$

$$\boldsymbol{\beta}_{\mathbf{i}} : \; g(\boldsymbol{\beta} \,|\, \mathbf{b}, \mathbf{W})$$

where
b is the mean of $\beta_{\mathbf{i}}$'s
W is the covariance of $\beta_{\mathbf{i}}$'s

Mixed Logit

Conditional on β_i the probability of choice is Logit

$$P(choice_{it} = j \mid \boldsymbol{\beta}_i) = P_{ijt}(\boldsymbol{\beta}_i) = \frac{e^{\boldsymbol{\beta}_i' \mathbf{X}_{ijt}}}{\sum_{k=1}^{J} e^{\boldsymbol{\beta}_i' \mathbf{X}_{ikt}}}$$

The likelihood of all choice made by customer i (conditioned on β_i) is:

$$L(\mathbf{Y}_i \mid \boldsymbol{\beta}_i) = \prod_{t=1}^{T_i} \prod_{j=1}^{J} \text{Prob}_{ijt}(\boldsymbol{\beta}_i)^{Y_{ijt}}$$

but each β_i is unknown

$$P(choice_{it} = j) = \int L_{ijt}(\boldsymbol{\beta}_i) g(\boldsymbol{\beta} \mid \mathbf{b}, \mathbf{W}) d\boldsymbol{\beta}$$ ← No closed form

Simulated Maximum Likelihood

Need to draw of a multivariate normal distribution

$\boldsymbol{\beta}_i \sim MVN[\mathbf{b}, \mathbf{W}]$;

Exploit the fact that $\mathbf{W}$ is a var-cov matrix and can be decomposed

$\mathbf{W} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$ where $\boldsymbol{\Gamma}$ = Upper triangular matrix

$\boldsymbol{\beta}_i = \mathbf{b} + \mathbf{w}_i$ where $\mathbf{b}$ and $\mathbf{w}_i$ are $K \times 1$ vectors

Make R standard normal draws of dimension K ($\mathbf{v}_i$)

$\boldsymbol{\beta}_i = \mathbf{b} + \mathbf{w}_i = \mathbf{b} + \boldsymbol{\Gamma}\mathbf{v}_i : \; MVN(\mathbf{b},\mathbf{W})$

Simulated Maximum Likelihood

- Make R draws for each customer – draws for each customer are the same across purchase occasions
  - Fix draws throughout the estimation
- Begin with initial guesses for **b** and Γ
- You can treat the R draws as S segment in latent class likelihood and use the 4 steps in the latent class case to find the maximum likelihood estimates for **b** and Γ

Bayesian Estimation

$$Posterior = \prod_{i=1}^{N} L(data \mid \boldsymbol{\beta}_{\mathbf{i}}, \mathbf{W}) \times priors$$

$$\begin{aligned} Prior &= N(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N \mid \mathbf{b}, \mathbf{W}) \ (normal) \\ &\times IW(\mathbf{W} \mid parameters) \ (Inverse\ Wishart) \\ &\times g(\mathbf{b} \mid assumed\ \ parameters) \ (Normal\ with\ large\ variance) \end{aligned}$$

Sequentially draw from the posteriors

$\mathbf{b} | \boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_N, \mathbf{W}$

$\mathbf{W} \mid \boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_N, \mathbf{b}$

$\boldsymbol{\beta}_{\mathbf{i}} \mid \mathbf{b}, \mathbf{W}$

Bayesian Estimation

$$p(\mathbf{b} \mid \boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N,\mathbf{W}) = Normal[\bar{\boldsymbol{\beta}},(1/N)\mathbf{W}]$$

$$\bar{\boldsymbol{\beta}} = (1/N)\sum_{i=1}^{N}\boldsymbol{\beta}_i$$

Easy to draw from MVN with known mean and Variance

$$p(\mathbf{W} \mid \mathbf{b},\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N) \sim Inverse\ Wishart[r + N, R + \sum_{i=1}^{N}(\boldsymbol{\beta}_i - \mathbf{b})(\boldsymbol{\beta}_i - \mathbf{b})']$$

Prior hyper-parameters

Bayesian Estimation

$$p(\beta_i \mid \mathbf{b},\mathbf{W}) : \ \underset{\text{Logit}}{L(data \mid \boldsymbol{\beta}_i)} \times \underset{\text{Normal}}{N(\boldsymbol{\beta}_i \mid \mathbf{b},\mathbf{W})}$$

No closed form – we can't simply sample from that distribution

Use Metropolis-Hastings Algorithm

Note that for Probit this leads to Normal posterior (conjugate prior)