37902 Foundations of Advanced Quantitative Marketing

Session III

1. Random Coefficients Models

1 Random coefficients models

Consider the model

$$U_{ijt} = \alpha_{ij} + X_{ijt}\beta_i + \epsilon_{ijt}$$

Thus far, we have allowed for α_{ij} and β_i to vary across households based on their demographic characteristics. The question we ask now is even without demographics, can we let these parameters vary across households? Think (as in a thought experiment) about two extreme situations:

1. We had only cross-sectional data, i.e., we had only 1 observation/household. In this case we can only have $\{\alpha_j, \beta\}$ and any heterogeneity has to come via demographics.

2. We had **lots** (!) of data on each household. In this case we could estimate a separate set of $\{\alpha_j, \beta\}$ for each household. So household 1 would have its own $\{\alpha_j, \beta\}$ and so would household 2, etc.

Fortunately, because we have panel data, we are not in situation 1. Unfortunately, we are not in situation 2. So we need an intermediate solution. What would that be? The answer is the random coefficients (RC) model.

RC model: Instead of estimating a separate set of parameters for each household (that we cannot). We will instead estimate the distribution of these parameters across households. Then we need only to estimate the parameter of the distribution conditional on some assumption we made about how the distribution looks like. This gives us the two most commonly used approaches.

1. Discrete heterogeneity distribution (or latent class/latent segment/finite mixture/semiparametric heterogeneity models)

2. Continuous heterogeneity distributions (continuous mixture model)

1.1 Discrete heterogeneity distribution

Here, we assume that $\Theta_i = \{\alpha_{ij}, j = 1, ..., J-1, \beta_i\}$ has a discrete distribution with a finite number of support points S. So $\Theta_i \sim D(\Theta_s, \Pi_s, s = 1, ..., S)$, where Θ_s is the value of Θ_i at support point s and π_s is the corresponding probability of belonging to that support s. Since it does not impose a specific discrete distribution on Θ_i, it is also called a semiparametric approach ("semi" because we are still assuming that the choice model is a logit). For a fully non-parametric approach, look at Briesch, Chintagunta & Matzkin (JASA, JBES). So how does this work? Let us start with the simplest case, i.e.,Θ_s can take two values Θ_1 and Θ_2. To distinguish it from other parameters, we will use superscripts instead, Θ^1 and Θ^2. IOW, we have:

$$\Theta^1 = \{\alpha_1^1, \alpha_2^1, \alpha_3^1, \beta_f^1, \beta_p^1\} \text{ for segment or class 1}$$

$$\Theta^2 = \{\alpha_1^2, \alpha_2^2, \alpha_3^2, \beta_f^2, \beta_p^2\} \text{ for segment or class } 2$$

$$\pi^1 = \frac{e^{\lambda_1}}{1 + e^{\lambda_1}}$$

$$\pi^2 = 1 - \pi^1$$

There will be 5+5+1=11 parameters in the case of the yogurt data. Since $\sum_{s=1}^{S} \pi^s = 1$ and $0 < \pi^s < 1$, the estimation task is to identify Θ^1, Θ^2 and λ_1 from the yogurt data. How do we do this?

Step 1 First, we revisit the data structure.

PanID	B_1	B_2	B_3	B_4	f_1	f_3	f_3	f_4	p_1	p_2	p_3	p_4	Params	P_c^1	P_c^2	L^1	L^2	L	$\ln(L)$
1													α_1	Step 3	Step 4			Step 6:	Step 7
1													α_2					$L_1 =$	
1													α_3					$L_1^1\pi^1 +$	
1													β_f					$L_1^2\pi^2$	
2													β_p						
2													α_1			$\underbrace{L_1^1,\ L_1^2}_{\text{Step 5}}$			
2													α_2						
2													α_3						
2													β_f						
2													β_p						
3																			
3													π_1						
3													Step 2: λ_1						
.																			
.																		Step 8:	$LL = \sum \ln(L_i)$

Step 2 We come up with initial values for the 11 unknown parameters. Note: π^1 is known as it is computed from λ_1.

Step 3 Compute the probability of the chosen brand assuming that the parameters come from segment 1. This is the column P_c^1 and you do it JUST LIKE FOR THE SIMPLE LOGIT MODEL YOU ALREADY ESTIMATED. For this you only use $\{\alpha_1^1,\ \alpha_2^1,\ \alpha_3^1,\ \beta_f^1,\ \beta_p^1\} = \Theta^1$.

Step 4 Repeat step 3 but assuming that the parameters come from segment 2. For this you only use $\{\alpha_1^2, \alpha_2^2, \alpha_3^2, \beta_f^2, \beta_p^2\} = \Theta^2$. This given the column P_c^2.

Step 5 Compute the likelihood for a household **conditional** on belonging to a segment. So take the product of P_c^1 for all the observations corresponding to household 1. Repeat this for P_c^2 (segment 2).

Step 6 Compute the **unconditional** likelihood for each household. So for household 1 compute $L_1 = L_1^1\pi^1 + L_1^2\pi^2$.

Step 7 Compute $\ln(L_i)$ for each household i.

Step 8 Add up the log-likelihoods from each household $LL = \sum \ln(L_i)$.

Step 9 Maximize LL to obtain Θ^1, Θ^2 and π^1.

2 LCM properties

Last session we started discussing random coefficients models and introduced the first type of RC model - the latent class/segment model or the discrete/semi-parametric heterogeneity model. Again keep in mind the key issues:

1) We believe that there is heterogeneity in preference and responsiveness to marketing activity that goes beyond being captured via demographic characteristics.

2) Data-wise, we are richer than having only cross-sectional data but poorer than having lots of observations for each panelist, i.e., we have moderately lived panel data.

Likelihood function for the LCM

$$P_{ijt|s} = \frac{exp(\alpha_{js} + X_{jt}\beta_s)}{\sum_{k=1}^{J} exp(\alpha_{ks} + X_{kt}\beta_s)}$$

$$\pi_s = \frac{exp(\lambda_s)}{1 + \sum_{l=1}^{S-1} exp(\lambda_l)}$$

i: Household, j: Brand, t: Purchase occasion, s : Segment.

A household's likelihood conditional on belonging to segment s is:

$$L_{i|s} = \left\{ \prod_{t=1}^{T_i} \prod_{j=1}^{J} P_{ijt|s}^{\delta_{ijt}} \right\}$$

$$\delta_{ijt} = \begin{cases} 1 & if\, i\, buys\, j\, at\, time\, t \\ 0 & otherwise \end{cases}$$

A household's unconditional likelihood:

$$L_i = \sum_{s=1}^{S} L_{i|s}\pi_s$$

Sample likelihood function is:

$$L = \prod_{i=1}^{N} L_i$$

So the sample log-likelihood function is:

$$LL = \sum_{i=1}^{N} ln \left(\sum_{s=1}^{S} \left\{ \prod_{t=1}^{T_i} \prod_{j=1}^{J} p_{ijt|s}^{\delta_{ijt}} \right\} \pi_s \right) \tag{1}$$

2.1 Some practical issues

Choosing starting values: a useful rule of thumb for choosing starting values is the following:

1. Fit one model without heterogeneity (unobserved) and return the 1-segment solution.

2. As starting values for the 2-segment solution, take the 1-segment values and perturb them up (by adding, say 0.01, to each parameter) and down (by subtracting, say 0.01, from each). Use these 2 vectors as the starting values and assign $\lambda_1 = 0$ (i.e.,$\pi_1 = \pi_2 = 0.5$).

3. Next, for the 3-segment solution, start with the 2-segment solution, append it with the 1-segment solution. Assign $\lambda_1 = \lambda_2 = 0$ (i.e., $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$) and proceed.

4. Continue the above procedure.

This will ensure that the likelihood will continue to show a marginal improvement when changing from S to $S + 1$ segments.

2.2 Using the estimates

2.2.1 Computing elasticity

There are two ways of thinking about elasticities from this model. The first is segment-level elasticity. Note that these are computed exactly the same way as for the simple logit model: calculate elasticities assuming all households belong to segment 1 and then conditional on them belonging to segment 2.

The second type of elasticity is the aggregate elasticity. This would be the market level elasticity. For this we first work out the unconditional probability (not likelihood) of household i buying brand j at occasion t as follows.

$$
\begin{aligned}
P_{ijt} &= \sum_{s}^{S} \pi_s P_{ijt|s} \\
\frac{\partial P_{ijt}}{\partial X_{jt}} &= \sum_{s=1}^{S} \pi_s \frac{\partial P_{ijt|s}}{\partial X_{jt}} \\
&= \sum_{s=1}^{S} \pi_s \beta_s P_{ijt|s}(1 - P_{ijt|s}) \\
e_{jj} &= \frac{\partial P_{ijt}}{\partial X_{jt}} \frac{X_{jt}}{P_{ijt}} = \frac{X_{jt} \sum_{s=1}^{S} \pi_s \beta_s P_{ijt|s}(1 - P_{ijt|s})}{\sum_{s=1}^{S} \pi_s P_{ijt|s}}
\end{aligned}
$$

The aggregate elasticity is simply the average of this elasticity across all purchase occasions for all houeholds. Similarly we can compute cross elasticity as well. Note from e_{jj} expression above that this is now free from the IIA property, i.e., they don't cancel off due to the $\sum$ sign. As a check if there is one 1 segment, then $e_{jj} = \frac{X_{jt}\beta P_{ijt}(1-P_{ijt})}{P_{ijt}} = \beta X_{jt}(1 - P_{ijt})$ as we had before.

2.2.2 Assigning households to segments

Intuition: If a household's likelihood computed with segment 1's parameters > likelihood computed with segment 2's parameters, the household is more likely to belong to segment 1. So we treat $\pi_s,\ s = 1, 2, 3, ..., S$ as "prior" probabilities of each household belonging to each segment. Then we "scale up" these prior probabilities by the corresponding likelihood to get a "posterior" that better reflects these likelihoods. IOW,

$$
\pi_{is} = \frac{\pi_s L_{i|s}}{\sum_{l=1}^{S} \pi_l L_{i|l}}
$$

Next, assign the household to the segment for which it has the highest posterior probability π_{is}.