37902 Foundations of Advanced Quantitative Marketing

Session V

Topics

1. Logit model with aggregated data
   (a) Role of outside good
   (b) Role of unobserved attribute
2. Heterogeneity

1 Logit model with aggregate data

What is aggregate data? It is typically store level (e.g., from Dominick's stores) or market level (e.g., Chicago market) data that represents the aggregation of all households in the marketplace or store area (refer back to class 1 notes). Recall the benefits (reflect all purchasing behavior) and the costs (difficult to understand loyalty, source of volume, etc.).

For example the coffee dataset on Google drive has market level data from LA market for the top 4 brands of coffee. The data contains the following at the weekly level (t = week)

1) Unit sales of coffee across all stores in LA for the 4 brands (j = brands)
2) Prices
3) Feature ads
4) Special displays
5) Feature and displays (when both are present)

3)∼5) are represented in ACV terms, so a number "20" under feature means stores that account for 20% of retail grocery sales in LA are feature advertising the brand in that week. A key feature of the data is that if I add up the sales across the four brands, they fluctuate over time, i.e., the total sales vary as a consequence of (a) seasonality(?), (b) prices of brands, (c)promotion of brands, and (d)other factors. Problem is in the logit model, summing the probabilities of brands always gives us 1. So prices do not influence demand for coffee according to our logit model. To deal with this we include a "no-purchase" option, i.e., sometimes, consumers do not buy coffee when they visit store. The utility of this option is usually given as

$$u_{i0t} = e_{i0t}, \text{ where } e_{i0t} \text{ is also Type-I EV distributed.}$$

In this case, when there is no heterogeneity,

$$P_{ijt} = \frac{exp(\alpha_j + X_{jt}\beta + \xi_{jt})}{1 + \sum_{k=1}^{J} exp(\alpha_k + X_{kt}\beta + \xi_{kt})}$$

Note: Since we now have $(J + 1)$ "brands", we can estimate J intercepts.

So

$$S_{jt} = P_{ijt}$$

Now we see if p_{jt} (price) of all brands goes down, then S_{ot} goes down as all the "inside shares" go up and the "outside good " share goes down. So what the model predicts is how shares change with prices and promotions. Where

$$S_{ot} = \frac{1}{1 + \sum_{k=1}^{J} exp(\alpha_k + X_{kt}\beta + \xi_{kt})}$$

$$S_{jt} = \frac{exp(\alpha_j + X_{jt}\beta + \xi_{jt})}{1 + \sum_{k=1}^{J} exp(\alpha_k + X_{kt}\beta + \xi_{kt})}$$

So how do I estimate α_j: $j = 1, ..., J$, β from aggregate data taking into account that $p_{jt} \in X_{jt}$ is correlated with ξ_{jt}?

Step 1 Transform unit sales into shares. How? For this we need the "sales" of the outside good or no-purchase option. This requires us to make an assumption on the potential market size each period, i.e., $Q_t = \sum_{j=1}^{J} Q_{jt} + Q_{0t}$, where Q_{jt} is the unit sales of brand j and Q_{0t} is the sales of the outside good. In the coffee spreadsheet I have assumed $Q_t = 1,000,000$ units (arbitrary). You can use other measures as discussed in class. Next compute $s_{jt} = \frac{Q_{jt}}{Q_t}$ and $s_{0t} = \frac{Q_t - \sum_{j=1}^{J} Q_{jt}}{Q_t}$.

Step 2 Compute the log odds ratio of the shares, i.e., $\ln(\frac{s_{jt}}{s_{ot}})$. Why?

I know that $\frac{s_{jt}}{s_{0t}} = exp(\alpha_j + X_{jt}\beta + \xi_{jt})$ and $\ln(\frac{s_{jt}}{s_{0t}}) = \alpha_j + X_{jt}\beta + \xi_{jt}$ or $Y_{jt} = \alpha_j + X_{jt}\beta + \xi_{jt}$.

This looks just like a linear regression! Although both s_{jt} and s_{ot} lie between 0 and 1, $\ln(\frac{s_{jt}}{s_{0t}})$ will lie between -∞ and ∞. So in principle ξ_{jt} can have full support as well. This is what makes the above equation amenable to a linear regression.

Step 3 Stack the data so we have 1 dependent variable column. So we have the following

Week	DV	Price	Feature	Display	F&D	α_1	α_2	α_3	α_4
1	$\ln(\frac{s_{11}}{s_{01}})$	p_{11}	f_{11}	d_{11}	fd_{11}	1	0	0	0
1	$\ln(\frac{s_{21}}{s_{01}})$	p_{21}	f_{21}	d_{21}	fd_{21}	0	1	0	0
1	$\ln(\frac{s_{31}}{s_{01}})$	p_{31}	f_{31}	d_{31}	fd_{31}	0	0	1	0
1	$\ln(\frac{s_{41}}{s_{01}})$	p_{41}	f_{41}	d_{41}	fd_{41}	0	0	0	1
2	$\ln(\frac{s_{12}}{s_{02}})$	p_{12}	f_{12}	d_{12}	fd_{12}	1	0	0	0
2	$\ln(\frac{s_{22}}{s_{02}})$	p_{22}	f_{22}	d_{22}	fd_{22}	0	1	0	0
2	$\ln(\frac{s_{32}}{s_{02}})$	p_{32}	f_{32}	d_{32}	fd_{32}	0	0	1	0
2	$\ln(\frac{s_{42}}{s_{02}})$	p_{42}	f_{42}	d_{42}	fd_{42}	0	0	0	1
...	...	...	...	...	...	...	...	...	...

Why do we do this? we want 1 β_p, 1 β_f, 1 β_d and 1 β_{fd}. So I need 1 column for each but I have 4 $\alpha's$ so I have to make sure that the right brand get the right α. To do this I create the 4 columns at the end of the above table. So in row 1, only α_1 will have an effect, in row 2 only α_2 will have an effect, etc.

Step 4 If p_{jt} and ξ_{jt} are uncorrelated I simply run OLS on the equation $Y_{jt} = \alpha_j + X_{jt}\beta + \xi_{jt}$ where ξ_{jt} is the error term. This gives me estimates of $\alpha_j,\ j = 1, ..., J$ and $\beta_p, \beta_f, \beta_{d,}\, and\, \beta_{fd}$.

Note: I need to set the overall intercept in the regressor = 0 since I cannot estimate 5 intercepts.

Step 5 What happens if p_{jt} and ξ_{jt} are not uncorrelated?

For this, let's go back to the first-order condition for price equation we had before.

$$p_{jt} = c_{jt} - \frac{1}{\beta_p(1 - P_{jt})}$$

As we said, c_{jt} is the cost faced by the firm. As researchers we observe some factors V_{jt} that affect costs and often factors η_{jt} that we do not.

So $c_{jt} = V_{jt}\psi + \eta_{jt.}$

$$p_{jt} = \underbrace{V_{jt}\psi}_{\text{part that does not contain } \xi_{jt}} - \underbrace{\frac{1}{\beta_p(1 - P_{jt})}}_{\text{part that contains } \xi_{jt}} + \eta_{jt}$$

The idea is the following. If I can "extract" the portion of p_{jt} that is uncorrelated with ξ_{jt} and use that in place of p_{jt} in Y_{jt} regression, then the "price" in that regression will be uncorrelated with the error term which will allow us to run OLS again. So how do we "extract" the right part of p_{jt}?

Step i Suppose I know V_{jt} then I can regress p_{jt} on V_{jt}. So $p_{jt} = V_{jt}\tilde{\psi} + \tilde{\eta}_{jt}$. $\tilde{\eta}_{jt}$ contains both η_{jt} and some function of of ξ_{jt}. Note: In this regression we also include other variables like intercepts, feature, display, f&d.

Step ii Compute the fitted value of price from the above regression. IOW,

$\hat{p}_{jt} = V_{jt}\hat{\tilde{\psi}}$

By construction, $\hat{p}_{jt}$ does not contain any "bad" stuff in it.

Step iii Replace p_{jt} in X_{jt} by $\hat{p}_{jt}$. This gives us $\hat{X}_{jt}$.

Step iv Run the regression using OLS. $Y_{jt} = \alpha_j + \hat{X}_{jt}\beta + \xi_{jt}$ ($\hat{X}_{jt}$ and ξ_{jt} are uncorrelated).

The above procedure is called "Two-stage least squares" or 2SLS. First stage - regress the "endogenous" variable (price) on the "instruments" (cost shifters V_{jt}) and other variables in X_{jt} that does not include p_{jt} (i.e., except p_{jt}). Second stage - regress the dependent variable ($Y_{jt} = \ln(\frac{s_{jt}}{s_{0t}})$) on the fitted value of prices in the first stage and all the other variables in X_{jt} (and the intercepts).

Issues: We need the right instruments - always a challenge. Because we cannot verify what is a good instrument and what is a bad one. Cost shifters usually work but may be "weak" (see Rossi's paper "Even the Rich Make Themselves Poor"). Second, standard errors in 2SLS are always larger than in OLS as we are using fitted (or estimated) values in the second stage.

2 Heterogeneity in the aggregate model with unobservable attributes

Before we take on 2 above, we first revisit using instruments with the aggregated logit model without heterogeneity.

In the previous session, we transformed the logit model into what is referred to as a logistic regression when we observe shares instead of 0 and 1 choices.

$$\ln\left(\frac{s_{jt}}{s_{ot}}\right) = \alpha_j + X_{jt}\beta + \xi_{jt}$$

We went through the steps to construct the data and then use OLS to get $\Theta = \{\alpha_j, j = 1, 2, ..., J, \beta\}$. Then we said that if X_{jt} and ξ_{jt} are correlated, we run 2SLS, first regress p_{jt} ($\in X_{jt}$) on the other elements of X_{jt} (not including p_{jt}), the intercepts (i.e., columns of 0/1), and the instruments (in this case cost shifters). Then get the fitted value of p_{jt}, i.e., $\hat{p}_{jt}$; replace p_{jt} in X_{jt} with $\hat{p}_{jt}$ and then run the above regression with $\hat{X}_{jt}$.

A potential downside to using the logit model is that it suffers from the IIA problem. So how to solve that problem?

As we saw with the LCM and Random Coefficients models the aggregate elasticities from these models are free from the IIA property although the individual level model is still logit and subject to IIA. By incorporating heterogeneity in α_j and β, we solve the IIA problem at the aggregate.

With aggregate data the problem is that we do not have purchase data at the individual level. Nevertheless, aggregate substitution patterns across brands are likely to deviate from the logit's IIA due to different consumers having different preferences. So if we could figure out a way to account for heterogeneity with the aggregate model, then we are in business, which means what identifies heterogeneity with aggregate data need not be heterogeneity per se (since we do not have individual data we can never verify this). Rather it is the deviation from IIA in the aggregate that helps us identify "heterogeneity" with aggregate data. This is important to keep in mind.

Now that we have clarified what we mean by heterogeneity with aggregate data, the real question is how do we incorporate it into the model? For this we will invoke the random coefficients model we saw before. According to this model,

$$P_{ijt} = \frac{exp(\alpha_{ij} + X_{jt}\beta_i + \xi_{jt})}{1 + \sum_{k=1}^{J} exp(\alpha_{ik} + X_{kt}\beta_i + \xi_{kt})}$$

$$\Theta_i = \{\alpha_{ij}, j = 1, ..., J, \beta_i\}$$

In the RC framework $\Theta_i \sim MVN(\Theta, \Sigma)$. So in week t, the aggregate share of brand j is obtained by "adding up" the above probabilities across the range over which the Θ_i's vary.

$$S_{jt} = \int \frac{exp(\alpha_{ij} + X_{jt}\beta_i + \xi_{jt})}{1 + \sum_{k=1}^{J} exp(\alpha_{ik} + X_{kt}\beta_i + \xi_{kt})} f(\Theta_i) d\Theta_i$$

What we observe are the shares s_{jt} and what our model predicts is S_{jt}. What we want to do is to minimize the distance between s_{jt} and S_{jt}, say by minimizing $\sum_{t=1}^{T}\sum_{j=1}^{J}(s_{jt} - S_{jt})^2$. The problem of course is the presence of the pesky ξ_{jt}. Since we do not observe it, we cannot compute S_{jt}.

So how do we solve this problem? First let us rewrite S_{jt} so as to separate the part that does not vary across households - call this δ_{jt} and the part that does vary across households - call this $\Delta\mu_{ijt}$. Here:

$$\delta_{jt} = \alpha_j + X_{jt}\beta + \xi_{jt}$$

and

$$\Delta\mu_{ijt} = \Delta\alpha_{ij} + X_{jt}\Delta\beta_i$$

Where $\Theta = \{\alpha_j, \beta, j = 1, ..., J\}$ is the mean of the MVN distribution of Θ_i; $\Delta\alpha_{ij}$ is individual $i's$ deviation of α_{ij} from α_j. IOW:

$$\alpha_{ij} = \alpha_j + \Delta\alpha_{ij}$$

and

$$\beta_i = \beta + \Delta\beta_i$$

Now $\Delta\Theta_i = \{\Delta\alpha_{ij}, j = 1, .., J, \Delta\beta_i\} \sim MVN(0, \Sigma)$.

$$S_{jt} = \int \frac{exp(\delta_{jt} + \Delta\mu_{ijt})}{1 + \sum_{k=1}^{J} exp(\underbrace{\delta_{kt}}_{\text{Part that does not depend on i}} + \underbrace{\Delta\mu_{ikt}}_{\text{Part that depends on i}})} f(\Delta\Theta_i) d\Delta\Theta_i$$

Notice that $\delta_{jt} = \alpha_j + X_{jt}\beta + \xi_{jt}$ looks a lot like the simple 2SLS regression we ran when we did not have heterogeneity. So, if we can isolate δ_{jt}, we can always run 2SLS. So how do we do that?

First we will leverage our knowledge of the RC model and use simulation to perform the above integration.

$$S_{jt} \approx \frac{1}{D} \sum_{d=1}^{D} \frac{exp(\delta_{jt} + \Delta\mu_{djt})}{1 + \sum_{k=1}^{J} exp(\delta_{kt} + \Delta\mu_{dkt})}$$

Where each draw d comes from $MVN \sim (0, \Sigma)$ - we already know how to make these draws. So here is our intuition in isolating and running 2SLS.

Aside: Remember, when we run OLS, we are minimizing $\xi'\xi$ (the sum of squared errors of ξ_{jt}'s) and in 2SLS we do the same but in the regression with fitted prices. So ultimately we pick parameters to make the $\xi's$ small.

Intuition:

1. Let us say, I told you Σ (or Γ where $\Gamma'\Gamma = \Sigma$). You can then make draws from this distribution to get $\Delta\Theta_d$. Since I know X_{jt}, I can compute $\Delta\mu_{djt} = \Delta\alpha_{dj} + X_{jt}\Delta\beta_d$.

2. Now looking at the share equation above. The only thing I do not know is $\delta_{jt}, j = 1, ..., J$. Now suppose I told you how to get δ_{jt} by "matching up" S_{jt} and s_{jt}. Then I would know δ_{jt} conditional on Γ.

3. Since I know δ_{jt}, I can now run the 2SLS as I did before but with δ_{jt} as dependent variable instead of $\ln(\frac{s_{jt}}{s_{0t}})$. This will give me a set of ξ_{jt}'s as residuals from this regression. These ξ_{jt}'s correspond to the specific Γ I chose through. So what do I do ?

4. I pick a different Γ and go through steps 1 - 3 again. This will give me a "new" set of ξ_{jt}'s. I now compare these ξ_{jt}'s with those I got from the first set of Γ parameters. How do I do that? I use a criterion similar to that I used in OLS ($\xi'\xi$) or 2SLS (some related quadratic form). If the new Γ yields smaller values of $\xi'\xi$, then I keep the new one and search for the next set of Γ to use.

5. When $\xi'\xi$ (or its equivalent criterion for 2SLS) does not change much between the "old" values of Γ and the new values of Γ, I stop. This gives me the "right" Γ.

6. I use the $\alpha_j, j = 1, .., J, \beta$ corresponding to the above Γ.

OK, so how do I do this in practice? Step 1 is the same as above. To formalize this, I write the intuition out as "pseudo-code" :

BEGIN Outer Loop (pick value of Γ)

BEGIN Inner Loop (compute δ_{jt} to match s_{jt} and S_{jt})
For the chosen Γ, compute $\delta_{jt}, j = 1, ..., J$

END Inner Loop

Run 2SLS: Compute ξ_{jt} as $\delta_{jt} - \hat{\alpha_j} - X_{jt}\hat{\beta}$
Compute quadratic form of ξ_{jt}

END Outer Loop

The key problem we face is the step in the inner loop - how to compute δ_{jt}/Γ to match s_{jt} and S_{jt}. This is the famous "contraction mapping" step. We will use this strategy in other models as well. The idea can be explained in the following steps.

Step 1: Start with some initial guesses for $\delta_{jt}, j = 1, 2, ..., J,\ t = 1, 2, ..., T$. So with 4 brands and 114 weeks, this will be a vector of 456×1 elements. We can pick all zeroes for δ_{jt} or use $\ln(\frac{s_{jt}}{s_{0t}})$ to approximate this. Call this vector $\delta_{jt}^{(0)}$.

Step 2 Given $\delta_{jt}^{(0)}$ and Γ, I can compute $S_{jt}^{(0)} = \frac{1}{D}\sum_{d=1}^{D} \frac{exp(\delta_{jt}^{(0)} + \Delta\mu_{djt})}{1+\sum_{k=1}^{J} exp(\delta_{kt}^{(0)} + \Delta\mu_{dkt})}$.

Step 3 Compute $\delta_{jt}^{(1)}$ as follows $\delta_{jt}^{(1)} = \delta_{jt}^{(0)} + \ln(s_{jt}) - \ln(S_{jt}^{(0)})$.

Step 4 Compute $C^{(1)} = \underbrace{max}_{t}(\underbrace{max}_{j}(absolute(\delta_{jt}^{(1)} - \delta_{jt}^{(0)})))$.

Step 5 If $C^{(1)}$ is very very small, then stop! You have gotten the "right" $\delta_{jt} = \delta_{jt}^{(1)}$.

Step 6 If $C^{(1)}$ is not small enough, go back to step 3, i.e., in general $\delta_{jt}^{(n+1)} = \delta_{jt}^{(n)} + \ln(s_{jt}) - \ln(S_{jt}^{(n)})$ till $\delta_{jt}^{(n+1)}$ and $\delta_{jt}^{(n)}$ are <u>very close.</u>

The actual criterion we minimize is not $\xi'\xi$ but $\xi' Z(Z'Z)^{-1}Z'\xi$, where Z is the matrix of instruments + exogenous variables in X + brand intercepts.