37902 Foundations of Advanced Quantitative Marketing

Session VI

Last week we went over the basic BLP model and how to estimate it with aggregate data. This session we will do the following:

1. Other approaches to accounting for endogeneity.
2. Extend what we have learned from
   (a) Household RC model estimation
   (b) Aggregat***ed*** data BLP estimation

   to account for endogeneity with household data.

1 The control function approach (See Petrin &Train, JMR)

Let's think about the two equations we care about

$$S_{jt} = \frac{1}{D}\sum_{d=1}^{D}\frac{exp(\alpha_{dj} + X_{jt}\beta_d + \xi_{jt})}{1+\sum_{k=1}^{J} exp(\alpha_{dk} + X_{kt}\beta_d + \xi_{kt})}$$

$$p_{jt} = Z_{jt}\gamma + \eta_{jt}, \text{ where } p_{jt} \in X_{jt}$$

The problem we face is that η_{jt} and ξ_{jt} are related in some complicated fashion unknown to us. The intuition behind the control function approach is as follows:

Intuition: Suppose I knew η_{jt}. I know ξ_{jt} is some complicated function of η_{jt} and vice versa. So I can write $\xi_{jt} = f(\eta_{jt})$ where $f(\cdot)$ is some flexible functional form, e.g. polynomial such as $\xi_{jt} = a_j\eta_{jt} + b_j\eta_{jt}^2 + c_j\eta_{jt}^3 + ...$ where $a_j, b_j, c_j...$ are brand j specific coefficients to be estimated. Now I can substitute $f(\eta_{jt})$ instead of ξ_{jt} in the share equation to obtain:

$$S_{jt} = \frac{1}{D}\sum_{d=1}^{D}\frac{exp(\alpha_{dj} + X_{jt}\beta_d + f(\eta_{jt}))}{1+\sum_{k=1}^{J} exp(\alpha_{dk} + X_{kt}\beta_d + f(\eta_{kt}))}$$

The above share equation has n unobserved error terms! So how do I estimate the parameters of the model?

Step 1: Run the auxiliary regression $p_{jt} = Z_{jt}\gamma + \eta_{jt}$ for brands $j = 1, 2, ..., J$ and compute $\hat{\eta}_{jt} = p_{jt} - Z_{jt}\hat{\gamma}$.

Step 2: Compute the "approximate" number of polynomials for η_{jt} (usually 1~3).

Step 3: Think of these η_{jt} terms as additional elements of X_{jt} for each brand j.

Step 4: Pick starting values of $\alpha_j, \beta, a_j, b_j, c_j, j = 1, ..., J$ (Assuming we are runing a cubic approximation) and of $\Gamma(\Sigma = \Gamma'\Gamma)$. Make draws from $MVN(\Theta, \Sigma)$ where $\Theta = \{\alpha_j, \beta, j = 1, ..., J\}$.

Step 5: Compute S_{jt} from the above equation.

Step 6: Iterate over the values of $\alpha_j, \beta, a_j, b_j,$ and c_j till $S_{jt} \approx s_{jt}$, say using non-linear least squares, i.e., minimizing $\sum_{t=1}^{T}\sum_{j=1}^{J}(S_{jt} - s_{jt})^2$.

2 The likelihood approach (see Park & Gupta JMR)

Let us focus on the same two equations as the control function approach. In addition we make the assumption that

$$\begin{pmatrix} \xi_{jt} \\ \eta_{jt} \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & \sigma_{\xi\eta} \\ \sigma_{\xi\eta} & \sigma_\eta^2 \end{pmatrix} \right)$$

So we are making the explicit assumption about the joint normality of the error terms in the two equations. Knowing the Cholesky trick from before, I can write the above equation in terms of 2 univariate standard normal draws ω_{1jt} and ω_{2jt} as follows.

$$\begin{pmatrix} \xi_{jt} \\ \eta_{jt} \end{pmatrix} = \begin{pmatrix} b_{11j} & b_{12j} \\ 0 & b_{22j} \end{pmatrix} \begin{pmatrix} \omega_{1jt} \\ \omega_{2jt} \end{pmatrix} = \begin{pmatrix} b_{11j}\omega_{1jt} + b_{12j}\omega_{2jt} \\ b_{22j}\omega_{2jt} \end{pmatrix}$$

So

$$\xi_{jt} = b_{11j}\omega_{1jt} + b_{12j}\omega_{2jt}$$

$$\eta_{jt} = b_{22j}\omega_{2jt}$$

$$\text{or}$$

$$\frac{1}{b_{22j}}\eta_{jt} = \omega_{2jt}$$

IOW

$$\xi_{jt} = b_{11j}\omega_{1jt} + \frac{b_{12j}}{b_{22j}}\eta_{jt}$$

From the control function approach I know $\eta_{jt} = p_{jt} - Z_{jt}\hat{\gamma}$ which can be computed from the auxiliary regression. So

$$\xi_{jt} = b_{11j}\omega_{1jt} + \frac{b_{12j}}{b_{22j}}(p_{jt} - Z_{jt}\hat{\gamma})$$

As with the control function approach, I can replace ξ_{jt} in the S_{jt} equation with the above equation. The only difference is that I now have an additional iid standard normal error component ω_{ijt} from which I need to make draws in addition to the MVN draws for Θ_i. The only difference is that these draws are brand and time specific.

Step 1: Run the auxiliary regression (step 1 of control function).

Step 2: Pick starting values for α_j, β, Γ and b_{11j}, b_{12j} and b_{22j}. It might appear that b_{12j} and b_{22j} are not separately identifiable but as we see below b_{22} can be identified from the likelihood for η_{jt}.

Step 3: Compute S_{jt} from the equation shown in the control function approach.

Step 4: Compute the joint likelihood of S_{jt} and p_{jt} as follows

$$\left\{ \prod_{t=1}^{T} \prod_{j=0}^{J} S_{jt}^{n_{jt}} \right\} \left\{ \prod_{t=1}^{T} \prod_{j=1}^{J} \frac{1}{b_{22j}\sqrt{2\pi}} e^{\frac{-(p_{jt} - Z_{jt}\gamma)^2}{2b_{22j}^2}} \right\}$$

n_{jt} in the above equation is the number of units of brand j sold in week t.

Step5: Iterate over the values of α_j, β, Γ, b_{11j}, b_{12j}, b_{22j} so as to maximize the joint likelihood in step 4.

Note: The Park & Gupta only works if n_{jt} is not large, i.e., there is sampling error associated with the computation of the shares s_{jt}. If there is no sampling error then the method breaks down (why?).

3 Accounting for endogeneity with household level data

With such data, several options are available to deal with the unobserved attribute ξ_{jt}. The only difference is that we now observed individual purchases (as opposed to aggregated shares). What are the methods that directly carry over from aggregated data analysis?

First, any likelihood based approach should be amenable to being used with household data since ultimately we will be doing MLE to estimate the model parameters, recall our estimation of i) the simple logit model; ii) the latent class model and iii) the continuous RC model. What does this mean? Since the Park & Gupta approach is also likelihood based, it appears that this would be a natural candidate for such an approach to account for endogeneity.

Next, the control function approach is agnostic to the procedure used in the "outer loop". So it can be used with aggregate data as well as with household level data. Again this method translates directly to use with household data. In the first stage run the auxiliary regression as with the aggregate data analysis, next construct the control function - i.e., polynomial of the residuals from the regression. Finally introduce the function of the error terms in the auxiliary regression as additional "X_{jt}" variables in the simple, latent class and RC models. This would address the endogeneity problem.

What are some other approaches we may use to account for endogeneity with household data?

3.1 The "brute-force approach"

This approach requires us to observe a "large" number of consumers purchasing each brand in each time unit of interest, e.g., a week. The basic idea here is that we are able to create brand-week shares much along the line of the aggregate data that we used previously to estimate "BLP-type" models.

Let's go back to the BLP model where at the <u>household level</u>, instead of aggregate shares we now have individual choices.

$$P_{ijt} = \frac{exp(\delta_{jt} + \Delta\mu_{ijt})}{1 + \sum_{k=1}^{J} exp(\delta_{kt} + \Delta\mu_{ikt})}$$

where

$$\delta_{jt} = \alpha_j + X_{jt}\beta + \xi_{jt}$$
$$\Delta\mu_{ijt} = \Delta\alpha_{ij} + X_{jt}\Delta\beta_i$$
$$\beta_i = \beta + \Delta\beta_i$$
$$\alpha_{ij} = \alpha_j + \Delta\alpha_{ij}$$
$$\text{and } \Delta\Theta_i = (\Delta\alpha_{ij}, j = 1, ..., J, \Delta\beta_i) \sim MVN(0, \Sigma)$$

In the RC model we estimated using household data, we had $\delta_{jt} = \alpha_j + X_{jt}\beta$. So the parameters to be estimated were $\alpha_j, j = 1, ...J, \beta, \Sigma$. In this case by contrast we also have ξ_{jt}, so we instead estimate $\delta_{jt}, j = 1, 2, ...J,\ t = 1, 2, ..., T$. And the parameters of Σ (or Γ to be more precise). So we estimate $\{\delta_{jt}, j = 1, ..., J,\ t = 1, ..., T,\ \Gamma\}$. δ_{jt} will be identified off the shares of each brand j in each week t. This is the reason we need to have a "large" number of observations for each brand-week. This will give us "reasonable" share estimates and hence estimates of δ_{jt}. The steps in the estimation are:

Step 0: Make D draws for each brand $j = 1, 2, ..., J$ for each household N.

Step 1: Pick starting values for δ_{jt} and Γ. For δ_{jt} a "smart" choice of initial values would be $\ln(\frac{s_{jt}}{s_{0t}})$ where s_{jt} is the share of brand j obtained by aggregating choices of brand j in each week t. So $n_{jt} = \sum_{i=1}^{N} I_{ijt}$ where $I_{ijt} = 1$ if household i buys j in week t, $j = 0, 1, 2, ..., J$. And $s_{jt} = \frac{n_{jt}}{\sum_{k=0}^{J} n_{kt}}$ (important to include "no-purchases" in the computation).

Step 2: As in the household RC models, compute each household's probability of the chosen brand for each draw d.

$$P_{ict|d} = \prod_{j=0}^{J} \left[\frac{exp(\delta_{jt} + \Delta\mu_{djt})}{1 + \sum_{k=1}^{J} exp(\delta_{kt} + \Delta\mu_{dkt})} \right]^{I_{ijt}}$$

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

$$\delta_{0t} = \Delta\mu_{d0t} = 0$$

Step 3: Construct each household's likelihood conditional on draw d, so $L_{i|d} = \prod_{t=1}^{T_i} P_{ict|d}$

Step 4: Construct the sample likelihood $L = \prod_{i=1}^{N} \frac{1}{D} \sum_{i=1}^{D} \prod_{t=1}^{T_i} P_{ict|d}$ and log-likelihood $\ln L = \ln(L)$

Step 5: Maximize $\ln L$ by iterating over the parameters δ_{jt} , $j = 1, ..., J$ and Γ.

Step 6: Run the 2SLS regression of δ_{jt} on exogenous variables and instruments.

The benefit of this approach is that you can use exactly the same code for this as you did for the RC logit model. The downside is that it is very time intensive and you could get large sampling error among the estimated δ_{jt}'s that will need to be incorporated when doing the 2SLS. To reduce the computational burden, we will leverage the contraction mapping step to estimate the parameters of this model.

3.2 Using the contraction mapping trick within the household likelihood function

Intuition: Recall that we said we observed a large number of households purchasing each brand including the outside good each period. This we can compute $s_{jt}, j = 0, ..., J$ for each period. Since we can compute s_{jt}, we will use knowledge of s_{jt} within the household's likelihood function to "invert" s_{jt} to obtain δ_{jt} instead of having to estimate the massive number of δ_{jt} parameters as part of the likelihood function as we had to do in the case of the "brute force" method.

Step 0: Same as the brute force method.

Step 1: (Outer loop) Pick starting values <u>only for the</u> $\Gamma(\Gamma'\Gamma = \Sigma)$ matrix. We will <u>only</u> estimate these parameters with the likelihood function.

Step 2: (Inner Loop) Pick starting values for the $\delta_{jt}{}'s = \delta_{jt}^{(0)}$.

Step 3: With knowledge of δ_{jt} and Γ compute[1] the following:

(3a) $P_{ijt|d}^{(0)} = \frac{exp(\delta_{jt}^{(0)} + \Delta\mu_{djt})}{1+\sum_{k=1}^{J} exp(\delta_{kt}^{(0)} + \Delta\mu_{dkt})}$

(3b) Next compute $P_{ijt}^{(0)} = \frac{1}{D} \sum_{d=1}^{D} \frac{exp(\delta_{jt}^{(0)} + \Delta\mu_{djt})}{1+\sum_{k=1}^{J} exp(\delta_{kt}^{(0)} + \Delta\mu_{dkt})}$

(3c) Now compute $S_{jt}^{(0)} = \frac{1}{\sum_{i=1}^{N} I_{ijt}} \sum_{i=1}^{N} I_{ijt} P_{ijt}^{(0)}$

(3d) Now compute $\delta_{jt}^{(1)} = \delta_{jt}^{(0)} + \ln(s_{jt}) - \ln(S_{jt}^{(0)})$

(3e) Iterate steps (3a)∼(3d) till the contraction map $\delta_{jt}^{(n+1)} = \delta_{jt}^{(n)} + \ln(s_{jt}) - \ln(S_{jt}^{(n)})$ converges. This gives us the "right" δ_{jt} for the choice of Γ in step 2, and concludes the inner loop.

[1] if you are making household specific draws, this will be $\Delta\mu_{dijt}$ as draws are i-specific.

Step 4: Now with the converged δ_{jt} and choice of Γ, construct the household likelihood function[2], as with the RC model, conditional on draw d.

$$L_{i|d} = \prod_{t=1}^{T_i} \prod_{j=0}^{J} P_{ijt|d}^{I_{ijt}}$$

Step 5: Next, compute the unconditional household likelihood $L_i = \frac{1}{D} \sum_{d=1}^{D} L_{i|d}$.

Step 6: Construct the sample likelihood $L = \prod_{i=1}^{N} L_i$.

Step 7: Pick values of Γ to maximize $\ln L$. This ends the "outer loop".

Step 8: Now with the values of δ_{jt} at the converged values of Γ, you can run 2SLS to recover $\alpha_j, j = 1, ..., J$ and β. This concludes the estimation.

[2]You can "read off" $P_{ijt|d}$ from the converged values of the inner loop.