ORGANIZACIÓN INDUSTRIAL EMPÍRICA IN7E0

Carlos Noton

Clase 5 - Martes 14 de Agosto
Primavera - 2018

Outline

Introduction

- We have introduced the discrete choice models as an alternative to the simple linear demand system that suffers from the *too many parameters problem*.
- Instead, using the consumer maximization problem and certain functional forms, elasticities can be constructed using fewer parameters.
- A terrible implication of the iid assumption in the logit model is the Independence of Irrelevant Alternatives (IIA) that has no economic background.
- We need to allow correlation for the idiosyncratic shocks of similar products in order to have more realistic patterns of substitution.
- We introduce the Random Coefficient Model that addressed most of the concerns about IIA while using aggregated market shares.

Mixed Logit or Random Coefficient Model a la BLP

Mainly developed by Berry, Levinsohn and Pakes (1995) and Berry (1994).

The standard model considers the following linear utility of consumer i for the model j in market t :

$$U_{ijt} = \alpha_i(y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}$$

where y_i is consumer's income, p_{jt} is the price and x_{jt} is the row vector of K observable characteristics, ξ_{jt} is an unobserved scalar product characteristic, and ε_{ijt} is a homoscedastic mean-zero stochastic term.

Same as before but now HETEROGENOUS CONSUMERS!!! Notice the subscripts (α_i, β_i)!!.

Random Coefficients

Notice the subscript i in the coefficients:

$$U_{ijt} = \alpha_i(y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}$$

α_i is the random coefficient that is individual specific and represents consumer i's marginal utility of income.

The marginal utility parameter vary across consumers but not across products for given a individual.

Random Coefficients

The distribution of the idiosyncratic parameter α_i is given by:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Gamma w_i + \Sigma v_i \qquad \text{where} \qquad v_i \sim \mathcal{N}(0_{K+1}, Id_{K+1})$$

where v_i is distributed as a standard normal shock and captures the unobservable consumer heterogeneity in price sensitivity. Also it could include some demographics w_i. For simplicity assume $\Gamma = 0$.

Although every individual has a different draw of coefficients (α_i, β_i), we will estimate the unknown parameters from the parametric distribution (not individual coefficients).

Define $\theta = (\alpha, \beta, \Sigma)$ as the vector containing all the parameters of the model.

Predicted Market Shares

The next step is to build the market shares consistent with this framework.

Market share s_{jt} of the product j is just an integral over the mass of consumers who choose model j (A_{jt}), that depends on random variables $\varepsilon = (\varepsilon_{i0t}, .., \varepsilon_{iJt})$ **and the individual shock v_i.**

Thus,

$$s_{jt}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \theta) = \int_{A_{jt}} dF_{\varepsilon}(\varepsilon|v_i)d\Phi(v_i) = \int_{A_{jt}} s_{ijt} d\Phi(v_i)$$

Predicted Individual Probabilities

The standard assumption is that ε is i.i.d.with Type I extreme value distribution, so we have a closed form for the individual probability s_{ijt}:

$$s_{ijt} = \frac{\exp(-\alpha_i p_{jt} + x_{jt}\beta_i + \xi_{jt})}{1 + \sum_h^J \exp(-\alpha_i p_{ht} + x_{ht}\beta_i + \xi_{ht})}$$

Same expression as before but now it includes individual coefficients (α_i, β_i).

Predicted Individual Probabilities

We replace (α_i, β_i) in terms of the common parameters (α, β, Σ):

$$-\alpha_i p_{jt} + x_{jt}\beta_i + \xi_{jt} \quad = \quad -\alpha p_{jt} + x_{jt}\beta + \xi_{jt} + [-p_{jt}, x_{jt}] \underbrace{\Sigma v_i}_{\widetilde{v}_i}$$

Notice that the shocks are not *iid* anymore. As long as Σ is not zero, $\widetilde{v}_i$ **allows for correlation associated to characteristics** $[-p_{jt}, x_{jt}]$.

For a given individual i, a larger α_i (or β_i) will be common to all products providing a larger rate of substitution between similar products (with similar $[-p_{jt}, x_{jt}]$).

Predicted Individual Probabilities

Using the expressions above, the individual probability is given by:

$$\begin{aligned} s_{ijt} &= \frac{\exp(-\alpha_i p_{jt} + x_{jt}\beta_i + \xi_{jt})}{1 + \sum_h^J \exp(-\alpha_i p_{ht} + x_{ht}\beta_i + \xi_{ht})} \\ &= \frac{\exp(-\alpha p_{jt} + x_{jt}\beta + \xi_{jt} + [-p_{jt}, x_{jt}]\Sigma v_i)}{1 + \sum_h^J \exp(-\alpha p_{ht} + x_{ht}\beta + \xi_{ht} + [-p_{ht}, x_{ht}]\Sigma v_i)} \end{aligned}$$

However, we only observed market shares aggregated at market level.

Market Shares based on Individual Probabilities

And the market shares are given by:

$$s_{jt}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \theta) = \int_{A_{jt}} \frac{\exp(-\alpha p_{jt} + x_{jt}\beta + \xi_{jt} + [-p_{jt}, x_{jt}]\Sigma v_i) d\Phi(v_i)}{1 + \sum_h^J \exp(-\alpha p_{ht} + x_{ht}\beta + \xi_{ht} + [-p_{ht}, x_{ht}]\Sigma v_i)}$$

How to compute this awful integral?

The non-analytical integral over the individual shocks v_i is computed through simulation (We'll see this soon).

The unobservable characteristic ξ_t is the only unobservable that explains an imperfect fit with the actual market shares, <u>playing the role of residual.</u>

Integral by Simulations

Suppose you need to compute the following integral:

$$s = \int g(x) dF(x)$$

There are several methods (quadrature, etc). Simulations are very simple. Choose a number of simulations R. Take R draws from F and evaluate the average of the draws evaluated in function g.

$$s = \frac{1}{R} \sum_{i=1}^{R} g(x_i)$$

Example 1 of Integral by Simulations

Suppose you need to compute the following single-dimensional integral:

$$s = \int x d\Phi(x) = \int x\phi(x)dx$$

Choose a number of simulations $R = 1000$. Take 1000 draws from a standard normal distribution $(x_1, x_2, .., x_{1000})$. Take the average of the terms:

$$\widehat{s} = \frac{1}{1000}\sum_{i=1}^{1000} x_i$$

This is just the estimator of the expected value of a standard normal, $E(x) = \int x\phi(x)dx$.

Can you see it? Does it make sense to you?

Example 2 of Integral by Simulations

Suppose you need to compute the following integral:

$$s = \int x^2 d\Phi(x) = \int x^2 \phi(x) dx$$

Choose a number of simulations $R = 1000$. Take 1000 draws from a standard normal distribution $(x_1, x_2, .., x_{1000})$. Get the 1000 values of each draw squared $(x_1^2, x_2^2, .., x_{1000}^2)$. Take the average of the squared terms:

$$\widehat{s} = \frac{1}{1000} \sum_{i=1}^{1000} x_i^2$$

Make the link with the variance. Compute $\int \tan(\log(x^6))\phi(x)dx$

Who needs primitive functions anymore!!

BLP Estimation

The estimates that rationalize the data is the vector $\widehat{\theta}$ such that:

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \| s_{jt}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \theta) - s_{jt} \|$$

However:

- The straightforward idea of matching predicted and observed market shares is unfeasible since ξ's are not observable and most variables enter in a non-linear fashion.
- To overcome this issue Berry (1994) and BLP (1995) developed an iterative process, in which the problem is linearized in ξ.

BLP Estimation

- After the problem is linearized in ξ, we follow a standard instrumental variables estimation to feed a GMM estimation.
- How to deal with the endogeneity problem? BLP suggested a set of instruments based on the characteristics of the competitors and within the same producer.
- Why these characteristics are good instruments? Based on models of differentiated products, markups are correlated with the distance of competitors in the product space. Are these characteristics correlated with ξ_{jt}?

Notation

Let us put together α and β inside of β to ease notation

So now $\beta_{K+1,1}$ is the vector of the $K+1$ linear parameters and Σ is the cholesky decomposition of the variance covariance matrix of the new β.

The variance-covariance matrix $\Sigma\Sigma'$ has dimension $(K+1) \times (K+1)$. The symmetric matrix have at most $(K+1)K/2$ parameters. Nevertheless, most covariances are set to zero for simplicity.

General Strategy

I summarize the three step procedure to estimate parameters as follows:

1. Given an initial value of Σ_0, find vector $\delta(\Sigma)$ of product-specific constants, called "mean utility" (**Simulation Stage**).
2. Use $\delta(\Sigma)$ to estimate $\beta(\delta(\Sigma)) = \beta(\Sigma)$ (**IV Stage**).
3. Compute GMM objective function $G(\Sigma, \delta, \beta) = G(\Sigma)$. Find $\widehat{\Sigma}$ that minimizes $G(\Sigma)$. (**GMM Stage**).

Simulation Stage: Finding deltas from Sigmas...

In order to find the "mean utility" vector $\delta(\Sigma)$, I need to compute the predicted market shares s_j for a given matrix Σ_0.

Recall that:

$$v \sim \mathcal{N}(0, Id) \Rightarrow \underbrace{\Sigma v}_{\widetilde{v}} \sim \mathcal{N}(0, \Sigma\Sigma')$$

Let us generate correlated shocks, $\widetilde{v}$, starting from *iid* shocks v.

Finding deltas from Sigmas...

We simulate R artificial consumers (that eventually can also include demographics).

So, let us draw shocks v from the K random coefficients for each of the R consumers:

$$\widetilde{v}_{K+1\times R} = \Sigma_{K+1\times K+1} \ \cdot \ v_{K+1\times R}$$

So we denote by $\widetilde{v}_i$ with $i = \{1, .., R\}$, the i^{th} column of $\widetilde{v}$ that is a $(K+1) \times 1$ vector of multivariate normal distribution with variance-covariance $\Sigma\Sigma'$.

Finding deltas from Sigmas to Predict Market Shares

Choose an arbitrary initial value for δ.
Compute the vector of predicted market shares, s, simulating the consumers as follows:

$$s_j(\Sigma, \delta) = \frac{1}{R} \sum_{i=1}^{R} \left[\frac{\exp(\delta_j + \widetilde{\widetilde{v}}_{ji})}{1 + \sum_{h=1}^{J} \exp(\delta_h + \widetilde{\widetilde{v}}_{hi})} \right]$$

where $\widetilde{\widetilde{v}}_{ji} = [-p_{jt}, x_{jt}]_{1 \times K+1} \widetilde{v}_i$, and $\widetilde{v}_i$ is a $(K+1)$ column vector with the taste shocks of individual i (the i^{th} column of $\widetilde{v}$)

Updating deltas

Strictly speaking to find an acceptable value of δ, I need to solve the J by J non linear system to match predicted, $s(\delta, \Sigma)$ and actual shares, s, for each market:

$$s(\delta, \Sigma) = s$$

for each given Σ.

Instead, BLP (1995) proved that we can use a recursive procedure to find the unique δ that ensures the predicted market shares to match the observed market shares (for a given Σ) (based on a Contraction mapping theorem.)

Contraction mapping theorem in BLP

Given a initial value of δ_0, the recursive formula to find the next round δ is given by:

$$\delta^{h+1} = \delta^h + \ln(s) - \ln(s(\Sigma, \delta^h))$$

Given an arbitrary small tolerance parameter, this procedure converges to the unique fixed point $\delta(\Sigma)$ that matches predicted and actual market shares.

It is very usual to start using the deltas consistent with the simplest logit: $\delta_0 = \ln(s/s_0)$, where s and s_0 are the vector of market shares of each product and the market share of the outside good respectively.

IV Stage : Finding betas

After obtaining the an acceptable $\delta(\Sigma)$, we now turn to estimate the vector $\beta(\Sigma)$. For that, we just need to run a simple instrumental variable regression as follows:

$$\delta(\Sigma) = X\beta + \xi$$

with the moment condition that $\mathbb{E}(Z'\xi) = 0$ for suitable instruments $Z_{N\times J}$ with $J \geq K+1$. Recall that X contains price, as is the extended matrix of characteristics.

IV Stage : Finding betas

The standard IV estimation lead us to:

$$\widehat{\beta}(\Sigma) = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'\delta(\Sigma)$$

where it is very usual to use the optimal weighting matrix $W_T = (Z'Z)^{-1}$, that minimizes the asymptotic variance.

GMM Stage: Updating Sigma through a GMM estimation

Given Σ, compute the residuals ξ as follows:

$$\xi(\Sigma) = \delta(\Sigma) - X\widehat{\beta}(\Sigma)$$

Hence, $\widehat{\Sigma}$ is:

$$\widehat{\Sigma} = \arg\min_{\Sigma \in \Theta} \xi(\Sigma)' Z (Z'Z)^{-1} Z' \xi(\Sigma)$$

where Θ is the set of feasible cholesky decompositions of a positive definite matrix.

Redo the procedure

Most optimization routines will start with Σ_0 and will update for a new $\widehat{\Sigma}$, and then you have to redo the same procedure until the change in the GMM objective function or the changes in $\widehat{\Sigma}$ are as small as you want.

It is crucial to keep the vector of draws, v, fixed during the estimation procedure to reach convergence. Still, a very complex optimization. Nevo (2001) shows the ugly formula for the standard deviations of the coefficients involved, which is a particular case of GMM.

Logit vs BLP

Logit Utility:

$$\begin{aligned} U_{ijt} &= \alpha(y_i - p_{jt}) + x_{jt}\beta + \xi_{jt} + \varepsilon_{ijt} \\ \text{IV Regression} &\Rightarrow \log\left(\frac{s_{jt}}{s_{0t}}\right) = -\alpha p_{jt} + x_{jt}\beta + \xi_{jt} \end{aligned}$$

Mixed Logit Utility:

$$\begin{aligned} U_{ijt} &= \alpha_i(y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt} \\ \text{Simulation +IV+ GMM} &\Rightarrow \min_{\theta\in\Theta} \|s_{jt}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \theta) - s_{jt}\| \end{aligned}$$

where s_{jt} is the observed market shares and $s_{jt}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \theta)$ are the predicted market shares:

$$s_{jt}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \theta) = \int_{A_{jt}} \frac{\exp(-\alpha p_{jt} + x_{jt}\beta + \xi_{jt} + [-p_{jt}, x_{jt}]\Sigma v_i) d\Phi(v_i)}{1 + \sum_h \exp(-\alpha p_{ht} + x_{ht}\beta + \xi_{ht} + [-p_{ht}, x_{ht}]\Sigma v_i)}$$

Logit vs BLP: Own price elasticities

Logit Own Price elasticities:

$$\begin{aligned} \frac{\partial s_{jt}(p_t)}{\partial p_{jt}} &= \frac{\partial s_{jt}(p_t)}{\partial \delta_{jt}} \frac{\partial \delta_{jt}}{\partial p_{jt}} = s_{jt}(1 - s_{jt})(-\alpha) \\ \Rightarrow \epsilon^L_{jj} &= \frac{\partial s_{jt}(p_t)}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = -\alpha p_{jt}(1 - s_{jt}) \end{aligned}$$

BLP Own Price Elasticities: BLP is a mixture of heterogenous logit consumers, hence it is not surprising that the own price elasticities are given by:

$$\epsilon^{BLP}_{jj} \equiv \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = -\frac{p_{jt}}{s_{jt}} \int |\alpha_i| s_{ijt}(1 - s_{ijt}) d\Phi(v_i)$$

Using our R simulated consumers, we can estimate the own price elasticity as follows:

$$\widehat{\epsilon}^{BLP}_{jj} = -\frac{p_{jt}}{s_{jt}} \frac{1}{R} \sum_{i=1}^{R} |\alpha_i| \widehat{s}_{ijt}(1 - \widehat{s}_{ijt})$$

Logit vs BLP: Cross price elasticities

Logit Cross price elasticities:

$$\frac{\partial s_{jt}(p_t)}{\partial p_{kt}} = \frac{\partial s_{jt}(p_t)}{\partial \delta_{kt}} \frac{\partial \delta_{kt}}{\partial p_{kt}} = s_{jt} s_{kt} \alpha$$

$$\Rightarrow \epsilon^{L}_{jk} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \alpha p_{kt} s_{kt}$$

Mixed logits or BLP Cross Price Elasticities:

$$\epsilon^{BLP}_{jk} \equiv \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} \quad = \quad -\frac{p_{kt}}{s_{jt}} \int |\alpha_i| s_{ijt} s_{ikt} d\Phi(v_i)$$

Obviously: $\widehat{\epsilon}^{BLP}_{jj} = -\frac{p_{kt}}{s_{jt}} \frac{1}{R} \sum_{i=1}^{R} |\alpha_i| \widehat{s}_{ijt} \widehat{s}_{ikt}$

Why IIA is less problematic in the BLP framework?

Conclusions

- We have covered the famous BLP model to estimate a demand for differentiated products. Huge improvement in comparison with previous models.
- Still, it has several limitations as it is static and characteristics are required to be exogenous.
- Next class will cover applications that include the supply side of the market to compute welfare and counterfactual exercises.