Advanced Industrial Organization II Problem Set 2

Joonhwi Joo*

Due: 2/11 (Th), in class

1 Random Coefficients Logit Model: Setup and Data

1.1 Specification

Individual i's utility of choosing product j is specified as:

$$\begin{aligned} u_{i,j} &= \alpha_i (w_i - p_j) + \mathbf{x}_j' \boldsymbol{\beta}_i + \xi_j + \epsilon_{i,j} \qquad \epsilon_{i,j} \sim i.i.d.\ T1EV \\ \begin{pmatrix} \alpha_i \\ \boldsymbol{\beta}_i \end{pmatrix} &= \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\Pi} \mathbf{q}_i + \sqrt{\boldsymbol{\Sigma}} \mathbf{v}_i \\ &= \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Pi}_\alpha \\ \boldsymbol{\Pi}_\beta \end{pmatrix} \mathbf{q}_i + \begin{pmatrix} \boldsymbol{\Sigma}_\alpha \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \end{pmatrix} \begin{pmatrix} v_{\alpha,i} \\ \mathbf{v}_{\boldsymbol{\beta},i} \end{pmatrix}, \end{aligned}$$

where $\mathbf{q}_i$ is the demographic variables, $\mathbf{v}_i$ is an idiosyncratic shock. The matrix $\boldsymbol{\Pi}$ represents the correlation between demographics and the coefficients, the lower-triangular matrix $\sqrt{\boldsymbol{\Sigma}}$ the correlation between the idiosyncratic shocks $\mathbf{v}_i$ and the coefficients. It is assumed that the distribution of $\mathbf{q}_i$ and $\mathbf{v}_i$ are fully known. For simplicity, the idiosyncratic shocks $\mathbf{v}_i$ is assumed to follow $\mathcal{N}(\mathbf{0}, \mathbf{I})$ with

$$\sqrt{\boldsymbol{\Sigma}} = \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{LL} \end{pmatrix}$$

so that $\sqrt{\boldsymbol{\Sigma}} \mathbf{v}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ diagonal.

When we allow for the individual coefficients, the right-hand side of the market share equation

*joonhwi@uchicago.edu

should involve the expectation over i. Specifically,

$$\begin{aligned} s_j &\simeq E\left[\Pr\left(i \to j\right)\right] && (1.1) \\ &= \int_{\mathbf{v}_i}\int_{\mathbf{q}_i}\int_{\left\{\epsilon_{i,j} | u_{i,j} \geq u_{i,k} \forall k\right\}} dF\left(\epsilon_{i,j}\right) dF\left(\mathbf{q}_i\right) dF\left(\mathbf{v}_i\right) \\ &= \int_{\mathbf{v}_i}\int_{\mathbf{q}_i} \frac{\exp\left(-\alpha_i p_j + \mathbf{x}_j' \boldsymbol{\beta}_i + \xi_j\right)}{\sum_{k \in \mathcal{J}_t} \exp\left(-\alpha_i p_k + \mathbf{x}_k' \boldsymbol{\beta}_i + \xi_k\right)} dF\left(\mathbf{q}_i\right) dF\left(\mathbf{v}_i\right) && (1.2) \\ &\simeq \frac{1}{n_s}\sum_{i=1}^{n_s} \frac{\exp\left(-\alpha_i p_j + \mathbf{x}_j' \boldsymbol{\beta}_i + \xi_j\right)}{\sum_{k \in \mathcal{J}_t} \exp\left(-\alpha_i p_k + \mathbf{x}_k' \boldsymbol{\beta}_i + \xi_k\right)}. && (1.3) \end{aligned}$$

1.2 Data

You are given two data files, "dataset_cleaned.csv" and "population_data.csv."

dataset_cleaned.csv contains a time series of product prices, shares, column of ones, three product characteristics (x_1-x_3) with six instruments for the prices (z_1-z_6). Note that in our data a market (market identifier is denoted by "id") is defined by different time periods of a single geographical region. For each market, the outside share s_0 is precalculated for your convenience.

population_data.csv contains two demographic variables "log_income" and "age," with $n_s = 1000$ predrawn iid $\mathcal{N}\left(\mathbf{0}, \mathbf{I}\right)$ shocks that correspond to each product characteristics in the "dataset_cleaned.csv."

2 Estimation

2.1 Questions

1. Suppose you include the constant term (a column of 1) in your estimation, while the utility from the outside option is still normalized to zero, i.e. for all i, $-\alpha_i p_0 + \mathbf{x}_0' \boldsymbol{\beta}_i + \xi_0 = 0$. What is the interpretation of the coefficient $\beta^{(1)}$ on the constant term?

   From here, you may or may not include the constant term in your estimation. Better if you can do both.

2. Estimate the simple logit model, using z_1-z_6 as the instruments for the prices. Report the estimates. Calculate the optimal GMM weighting matrix from this simple logit model estimates.

3. What are the dimensions of the parameters to be estimated in the full model, i.e. $\left(\alpha, \boldsymbol{\beta}, \boldsymbol{\Pi}, \sqrt{\boldsymbol{\Sigma}}\right)$? (Note. the answer should be different upon your choice of including the constant term.)

4. Estimate the full model. Report the parameter estimates. Are the parameter estimates $\left(\hat{\alpha}, \hat{\boldsymbol{\beta}}\right)$ of the full model very different from what you have estimated in the simple logit

model?

(a) You are fine to fix the optimal weighting matrix calculated from Step 2 during the optimization.

(b) Use the estimates of the simple logit model as the starting value for the nonlinear optimization first. If possible, also try different starting values afterwards.

2.2 Some Advices

- You may either use NFP or MPEC. A Matlab/Knitro MPEC code by Dube, Fox, and Su (2012) is also attached for your reference. However, their code does not contain the estimation of $\mathbf{\Pi}$. In order to estimate $\mathbf{\Pi}$ you will have to redo many parts of the code over again anyway. (And very often writing your own code from the beginning is easier than modifying the code of someone else.)

- You are free to use any computer language and any reliable optimizer (including KNITRO, IPOPT, but not fmincon). A general advice is not to use interpreted languages (such as R or Python[1]) because they can be too slow for the problem to be feasible.

- Read the online appendix of Dube, Fox, and Su (2012) carefully. Along with other helpful information, it contains most of the closed-form first and second order derivatives that are required to solve the problem.

- You will have to supply the sparsity patterns of the Jacobian of the constraints and Hessian of the Lagrangian to the optimizer.

[1]Unless you attach CPP codes in R, or use Numpy or Scipy in Python.