Econ 7218 Problem Set 2

Due by Monday, May 8, 2023

Question I. For this problem set you will need to download the dataset constructed by Ciliberto and Tamer (2009), which is available from the *Econometrica* website at this URL. (The same data is also available on NTU COOL). Using this dataset to replicate as closely as possible the following specification of Berry (1992):

1. Ordered probit with no observed heterogeneity:

$$\pi_i(N) = X_i\beta - \delta \ln(N) + u_{i0},$$

   where u_{i0} is iid $N(0,1)$.

2. Simulation estimation (MSM or MSL) with

$$\pi_{ik}(N) = X_i\beta - \delta \ln(N) + Z_{ik}\alpha + \sigma u_{ik} + \rho u_{i0},$$

   where u_{ik} and u_{i0} are iid $N(0,1)$ and $\sigma = \sqrt{1-\rho^2}$. Use the equilibrium selection rule where the most profitable firms move first.

For the market-specific variables X_i, use a constant, population, and distance. For the firm-specific variables Z_{ik} in the second specification above, use market presence and cost (distance to the airline's hub).

Write up your results as a short section of an empirical paper. Describe the data, the model specifications, and interpret your results. For each specification, report the estimated coefficients β, δ, α, and ρ with standard errors. Provide the relevant details of the computational aspects behind the results.

Question II. EM algorithm for estimating Gaussian mixture model.

- Draw 2,000 points from $\mathcal{N}(20, 5^2)$ and 3,000 points from $\mathcal{N}(10, 3^2)$ and mix them together. Plot a histogram.

- Denote parameters $\mu = [\mu_1, \mu_2]$, $\sigma^2 = [\sigma_1^2, \sigma_2^2]$, and $\pi = [\pi_1, \pi_2]$, the log likelihood can be expressed as

$$\ell(x; z, \mu, \sigma^2, \pi) = \sum_{i=1}^{n} \sum_{r=1}^{2} I(z_i = r)[\ln \pi_r + \ln f(x_i; \mu_r, \sigma_r^2)],$$

where $z = [z_1, \cdots, z_n]$ denotes the component indicator. In the E step, the conditional probability to which component each observation x_i belongs can be computed by

$$Pr(z_i = r|x_i; \mu, \sigma^2, \pi) = \frac{\pi_r f(x_i; \mu_r, \sigma_r^2)}{\sum_{r=1}^{2} \pi_r f(x_i; \mu_r, \sigma_r^2)},$$

then the expected log likelihood can be expressed as

$$Q = \sum_{i=1}^{n} \sum_{r=1}^{2} Pr(z_i = r|x_i; \mu, \sigma^2, \pi)[\ln \pi_r + \ln f(x_i; \mu, \sigma^2)].$$

In the M step, the expected log likelihood is maximized with respect to parameters, which can be expressed as follows:

$$\pi_r = \frac{\sum_{i=1}^{n} Pr(z_i = r|x_i; \mu, \sigma^2, \pi)}{n}$$

$$\mu_r = \frac{\sum_{i=1}^{n} x_i Pr(z_i = r|x_i; \mu, \sigma^2, \pi)}{\sum_{i=1}^{n} Pr(z_i = r|x_i; \mu, \sigma^2, \pi)}$$

$$\sigma_r^2 = \frac{\sum_{i=1}^{n} (x_i - \mu_r)^2 Pr(z_i = r|x_i; \mu, \sigma^2, \pi)}{\sum_{i=1}^{n} Pr(z_i = r|x_i; \mu, \sigma^2, \pi)}$$

References

Berry, Steven T (1992) "Estimation of a Model of Entry in the Airline Industry," *Econometrica: Journal of the Econometric Society*, pp. 889–917.

Ciliberto, Federico and Elie Tamer (2009) "Market structure and multiple equilibria in airline markets," *Econometrica*, Vol. 77, pp. 1791–1828.