EC8559, Problem Set 2: Logit, Nested Logit, and Random-Coefficient Logit

Julie Holland Mortimer

Due: 12.30pm ET, Wednesday, November 1, 2023

Instructions As in your first problem set: when asked to describe an estimation algorithm, please provide enough detail so that an RA who knows Matlab but has never taken any IO could use your description to write the estimation program. Actual estimation can be done in groups if you prefer. Please provide individual write-ups of your work, and note the members of your group. Attach a printout of your programs to your solutions.

Data Download "data_ps2_3nests.txt" from the course's web page. It is a simulated panel dataset that contains the following information:

Column	1	2	3	4	5	6	7	8	9
Variable	car id	year	firm id	price	quantity	weight	hp	ac	nest3

Market size Assume throughout that the market size M is equal to 100 million in each year, which is approximately the number of households in the US.

Outside good Let the share of the outside good vary across the three years, but keep M fixed throughout.

Normalization and instruments Estimation procedures can have problems when they contain variables that differ by several orders of magnitude. To avoid problems you should normalize the covariates you use. You will also need to create instruments. For all estimation procedures in this problem set, create the instruments from the original values, and then normalize the continuous covariates before performing the regression. Do not normalize dummy variables, constants, the log of within group share (in the nested logit problem), the left-hand side variables (functions of deltas or shares), or the instruments that you create. In other words, you only need to normalize weight, horsepower, and price. Normalize by dividing each variable by its mean.

Now suppose that the model is a logit model. The difference between the vertical model and this logit model is that the utility error is now i.i.d. extreme value and $\alpha_i = \alpha$ for all i, i.e.:

$$\boldsymbol{u_{ijt} = \delta^*_{jt} + \epsilon_{ijt}} \quad \text{and} \quad \boldsymbol{\delta^*_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} \quad \forall\, t \in T}$$

1. Beginning with the utility function, derive market share as a function of δ^*s. Then invert the equation to solve for δ^* as a function of shares. Allow each year to be separately estimated.
2. If instead, I asked you to pool the data and estimate a single model, what would change in your derivation?
3. Estimate the demand system parameters using GMM with just the demand-side moment conditions. Since price is endogenous, you will need to use at least one instrument. Construct the BLP instruments (characteristics of competing products **in a given year**) analogously to the vertical demand problem set.
4. Calculate and report own- and cross-price elasticities for the ten highest-selling vehicles in the dataset using your estimates and explain why they may be unrealistic.

Now use a nested logit model with a single level of nesting.

5. (a) Write down market share for a single product as a function of the vector $\boldsymbol{\delta}^*$ and the nesting parameter σ. Use the σ and group notation used by Berry (1994 Rand), not the λ notation used by Train and McFadden.
   (b) Invert that equation to solve for δ_j^* as a function of market shares, within group shares, and σ.
   (c) Finally, use this second equation to write a regression equation with an observed quantity on the left-hand-side and observed variables and coefficients (including σ) on the right-hand-side.
6. (a) How is nested logit an improvement over plain logit?
   (b) Why might all forms of nested logit be problematic?
7. (a) Estimate the demand system parameters using just the demand-side moment conditions for this nested-logit model. Instrument for price using the average of product characteristics (i.e. weight, hp, ac) of products produced by other firms **in a given year**. Similarly, construct the instrumental variables by using the average of product characteristics of other products (including those of the same firm) within the same group in a given year. For your nests, group products based on the "nest3" variable. Include the outside good in its own additional group in each year. Calculate 2SLS estimates, allowing all coefficients to vary across the three years of data, and use them as starting values for the `fminsearch` command in your GMM routine. Report your results.
   (b) Could a researcher estimate a random-coefficient logit model using only one year of this dataset? Why or why not?
   (c) Calculate and report own- and cross-price elasticities for the ten highest-selling vehicles in the dataset using your estimates. How do they compare to your estimates in question 4?
   (d) Suppose instead that we estimated a version of a nested-logit model that pooled all three years' worth of data. What assumption on substitution patterns is implicit in this choice?

8. Suppose we were interested in improving the substitution patterns.
   (a) Would the Multinomial Probit model be appealing in this setting? Why, or why not?
   (b) Would the Pure Characteristics model be appealing in this setting? Why, or why not?

PyBLP is a Python package developed by Christopher Conlon and Jeff Gortmaker to estimate differentiated products demand systems and released in 2021. It is popular among economists, and we will use this problem set as an opportunity to familiarize ourselves with it. For the last question, we will follow Chris and Jeff's tutorial on how to estimate a random coefficients logit model using data on cereal consumption from Nevo (2000).

- Download the PyBLP package https://pyblp.readthedocs.io/en/stable/index.html
- Follow the random coefficients logit tutorial through "Setting Up and Solving The Problem Without Demographics": https://pyblp.readthedocs.io/en/stable/_notebooks/tutorial/nevo.html

9. Report the estimates of the random coefficients model using an unrestricted covariance matrix for random tastes using Monte Carlo integration.