Problem Set 1: BLP Demand Estimation

Allan Collard-Wexler
Daniel Xu

Due Date: October 12 (by midnight), 2022

The data from the problem set is OTC_Data.csv OTC_Headache.csv. This problem set uses data on "Over the counter" Headache medicine (i.e. aspirine, tylenol and such) and was graciously provided by Vishal Singh at NYU Marketing, so if you want good demand data make friends with Marketing people! The data is at the store/week level for 3 brands and 3 package sizes.

- Count: Number of People that go into the store each week.

- Promotion: Is there a promotion on the product that week.

- Price: Price of the package.

- Week and Store are the time and market indicator.

Demographic data for this problem set is composed of the following:

- Income*: Household Income for person #.

- Week and Store are the time and market indicator.

1 Logit

Consider the utility function for product j in store-week t for consumer i:

$$u_{ijt} = X_{jt}\beta + \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt} \qquad (1)$$

Brand	Size (tab)	Market Share	Unit Price	Price/50 tab	Unit Wholesale Price
Tylenol	25	8.9%	3.52	6.91	2.26
Tylenol	50	11.1%	4.99	4.99	3.76
Tylenol	100	7.6%	7.14	3.57	5.93
Advil	25	7.3%	3.02	6.29	2.02
Advil	50	5.1%	5.19	5.19	3.70
Advil	100	2.2%	8.23	4.12	6.25
Bayer	25	2.5%	2.66	5.35	1.90
Bayer	50	2.0%	3.81	3.81	2.67
Bayer	100	4.9%	4.05	2.02	3.73
Store Brand	50	6.2%	3.57	3.57	2.48
Store Brand	100	4.2%	3.29	1.65	1.51

Table 1: Summary Statistics for Headache Data

where ϵ_{ijt} is an i.i.d. logit draw, X_{jt} are observed product characteristics and ξ_{jt} are unobserved product characteristics.

Hint: Do this part in STATA or R, since it will be useful to check your MATLAB (or Python, Julia) code later on. As well, never ever perform data cleaning and manipulation in MATLAB or things like fixed-effect regressions. You often get wrong or inaccurate results if you don't know all the tricks that people use to do these things.

Estimate this model:

1. Using OLS with price and promotion as product characteristics.

2. Using OLS with price and promotion as product characteristics and brand dummies.

3. Using OLS with price and promotion as product characteristics and store-brand (the interaction of brand and store) dummies.

4. Estimate the models of 1, 2 and 3 using wholesale cost as an instrument.

5. Estimate the models of 1, 2 and 3 using the Hausman instrument (average price in other markets).

6. Using the analytic formula for elasticity of the logit model, compute the mean own-price elasticities for all brand in the market using the estimates in 1, 2 and 3. Do these results make sense? (Discuss)

- Working in groups on these numerical problem sets is fine, and encouraged. All members of a group should ultimately do the calculations and hand them in individually.

- When asked to report results present the answer in a table. Nothing fancy but don't simply attach a printout of the statistical program you used. A good habit is to mimic the tables in an AER paper. There is nothing worse than getting interesting results which are unreadable to anyone but you. You should attach the code you used to generate the results as an appendix.

2 Random-Coefficients Logit, a.k.a. BLP

Now consider the following model for demand with individual heterogeneity on coefficients:

$$u_{ijt} = X_{jt}\beta + \beta_{ib}B_{jt}(\text{Branded Product}) + \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \qquad (2)$$

The random coefficients are determined as follows:

- There is a random coefficient on branded products $\beta_{ib} = \sigma_B v_i$ where v_i is a draw from a standard normal distribution.

- The price coefficient depends on income: $\alpha_i = \alpha + \sigma_I I_i$ (Income).

1. Estimate the parameters of the model β, α, σ_{ib} and σ_I using BLP. As instruments, use both cost instruments and the prices of the same product in the same week at other stores, not just the average but a variable for each price at another store in the same period and choose 30 other stores. As a weight matrix, choose the optimal weighting matrix $(Z'Z)^{-1}$. I've put the instruments I use on the webpage in case you have trouble with this part.

   If you use the trick of separating linear and non-linear parameters then have brand dummies and promotion in your product characteristics. If you don't use this trick, just estimate α, σ_{ib} and σ_I.

   Use fminsearch in MATLAB to minimize the GMM criterion function (or a non-gradient based approach otherwise), and use both 0's as your

starting value and something else. Just give the point estimates and the GMM Criterion value. You don't have to compute standard errors for this problem.

2. What are the elasticities, both own-price and cross-price for store 9 in week 10? How are these different from the logit model (just set the σ's to 0 to see this). Discuss.

3. Back out the marginal costs for store 9 in week 10 under the assumption that each brand is owned by a single company. How different are these from the wholesale costs?

3   Merger Analysis

Suppose that Tylenol, Advil and Bayer merge.

1. Predict prices using the logit model (no random coefficients here) after the merger, but only for store 9 in week 10. Make sure that if there is no merger the prices you get don't change!

2. Write down how you would predict the change in prices after the merger using the random coefficient model. You don't have to actually perform this task, just tell me how you would do it.

3. Challenge Question: Predict prices using the random coefficients model you estimated after the merger. Show the effects for store 9 in week 10.