Problem Set 1: Demand Estimation
Econ 760: Empirical Industrial Organization

Matthew Grennan*

Due Date: Feb 12, 2025

General Notes:

- Talking to and helping each other with these problem sets is fine, and encouraged. I recommend trying to do everything yourself and giving yourself a chance to struggle a little before you seek help, though. Everyone 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.
- You should discuss and interpret all results. It should be easy for the reader to read and learn from your writeup. Economics is about communication and explanation as much as it is calculation.
- Use whatever statistical/optimization software you prefer. I come from the "use Stata until you have to use Matlab" generation, so I have solutions using that approach. You may prefer R, Python, or a different mix. You can also use the PyBLP API interface if you like. PyBLP is a great resource and the tutorials are super informative. Using the APIs will make much of the assignment much easier, but the tradeoff is that you won't learn the nitty gritty yourself, which you may find limits you in the future. I leave it to you to make the right tradeoffs for yourself.
- You should attach the code and/or description of the process you used to generate your results as an appendix.
- The Challenge Questions are all optional. But you think you want to do IO, you should at least give them a shot, even if you don't finish them.

The Data:
This problem set uses data on "over the counter" headache medicine (i.e. aspirin, Tylenol, etc.) and was graciously provided by Vishal Singh (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 as well as a generic store brand in 2 sizes.

Sales data OTC_Sales.csv

- Product ID: IDs 1-11 corresponding to brand names and package sizes as in the sum stats below. (Note data is in "wide" format, so these are subscripts on variables.)
- Store: Identifier for the store.
- Week: Identifier for the week.
- Count: Number of People that go into the store each week.

*This Problem Set was partially developed with Allan Collard-Wexler. Any errors are my own.

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

Demographic data OTC_Demographics.csv

- Store: Identifier for the store.
- Educ: Fraction of customers college educated.
- Income: Natural log of average customer household income.

The following table provides the mapping from product ids to brand names and bottle sizes:

Product ID	Brand Name	Size (tablets)	Market Share	Unit Price	Price/100 tablets	Unit Wholesale Price
1	Tylenol	25				
2	Tylenol	50				
3	Tylenol	100				
4	Advil	25				
5	Advil	50				
6	Advil	100				
7	Bayer	25				
8	Bayer	50				
9	Bayer	100				
10	Store Brand	50				
11	Store Brand	100				

Table 1: Summary Statistics for Headache Data

Assigned Questions:

1. **Sum Stats and Institutional Details**
   (a) Do some brief diligence on the products and industry. What do you anticipate may be some important determinants of demand, substitution, and pricing?
   (b) Complete the table above by adding columns for the mean: market share, unit price, price/100 tablets, and unit wholesale price. Interpret any notable patterns you see in the summary statistics.
2. **Logit**
   Consider the utility function for product j in store-week t for consumer i:

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

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

   Estimate this model:
   (a) Using OLS with price and promotion as product characteristics.
   (b) Using OLS with price and promotion as product characteristics and product fixed effects (where a "product" is a brand-size combination).
   (c) Using OLS with price and promotion as product characteristics and product-store (the interaction of product and store) fixed effects.
   (d) Estimate the models of (a), (b) and (c) using wholesale cost as an instrument.
   (e) Estimate the models of (a), (b) and (c) using the Hausman instrument (average price in other markets).

(f) Why can't we use BLP instruments in this setting?

(g) Using the analytic formula for elasticity of the logit model, compute the mean own-price elasticities for all products in the market using the estimates in (a), (b) and (c). Do these results make sense? (Discuss)

3. Random-Coefficients Logit Demand + Bertrand-Nash Supply (a.k.a. BLP)

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

$$u_{ijt} = X_{jt}\beta + \beta_i^B \mathbf{1}\{\text{Branded Product}\} + \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{3}$$

The random coefficients are determined as follows:

- There is a random coefficient on branded products $\beta_i^B = \sigma_B v_i$ where v_i is a draw from a standard normal distribution.
- The price coefficient depends on (log)income: $\alpha_i = \alpha + \sigma_I I_i$ (Income).

(a) Estimate the parameters of the model β, α, σ_B 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 posted the instruments I use in OTC_Instruments.csv in case you have trouble with this part.

If you use the trick of separating linear and non-linear parameters then include product fixed effects and promotion in your product characteristics. If you don't use this trick, just estimate α, σ_B and σ_I.

Use fminsearch in MATLAB to minimize the GMM criterion function, 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.

(b) 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.

(c) 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?

4. Merger Analysis

Suppose that Tylenol, Advil and Bayer merge.

(a) 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!

(b) 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.

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

5. Shortages and Demand

(a) Suppose there is a shortage of Tylenol, and all Tylenol products have been removed from the market. What will happen to the market share of the remaining products? i.e. Where are consumers likely to shift their demand? Predict shares with these shortages.

(b) Challenge Question: Now suppose that Tylenol and Bayer are not always available. When a consumer enters the store, there is a 50% chance that Tylenol is available and a 50% chance that Bayer is available. (Assume these two events are independent) Estimate the demand under these conditions.

6. Challenge Question: New Product Introduction

The chain of stores you have data from is considering introducing a 25 tablet size bottle.

(a) Given what you know about the institutional setting and data available here, write down what you think could be a better model of demand than those we have tried so far, remembering your goal is to capture important substitution patterns relevant to this product introduction. Argue for why you think this model should be better. Estimate your model and discuss the results.

(b) What will be the expected demand for the new product at a price of $2.00?

(c) What will the optimal price be (feel free to assume a reasonable number for costs and treat all products as pricing independently if you want to simplify things a bit, or do it "right" if you want to challenge yourself further)? What will the new market sum stats look like after this product introduction?