ECN 594: Homework 1

Demand Estimation

Due: See Canvas

You may work in groups of up to 2 people.

Instructions

- This assignment is out of 100 points, with 10 bonus points possible. The maximum score is 100/100 (so if you earn 110 points, you receive 100/100).
- You may work in groups of up to 2 people. If working in a group, please submit one assignment with both names.
- Submit your solutions as a PDF along with your Python code (either a `.py` file or Jupyter notebook).
- Use the `pyblp` package for estimation. Documentation is available at https://pyblp.readthedocs.io/.
- This assignment uses the same cereal dataset as in lecture. The data files are available on Canvas.

Question 0: Submission (5 points)

Submit all materials to Canvas by the deadline:

- A PDF containing your written answers to all questions
- Your Python code (either a `.py` file or Jupyter notebook `.ipynb`)
- All figures should be embedded in your PDF and clearly labeled
- If working in a group, include both names on your submission

Background and Data

In this homework you will estimate demand for breakfast cereals using the methods we covered in class. You will use the `pyblp` package in Python to estimate logit and logit with demographic interactions models.

There are two datasets:

- **product_data.csv**: Contains data about market shares, prices, and product characteristics.
  - *market_ids*: Market identifiers (city-quarter, e.g., 'C01Q1' = city 1, quarter 1)
  - *product_ids*: Product identifiers
  - *shares*: Market shares
  - *prices*: Prices
  - *sugar*: Sugar content (grams)
  - *demand_instruments0, ..., demand_instruments19*: Pre-computed demand instruments
- **agent_data.csv**: Contains consumer demographic data for each market.
  - *market_ids*: Market identifiers
  - *weights*: Weight of each consumer draw (= 1/20)
  - *income*: Draw from the income distribution in each market

Question 1: Basic Logit Model (45 points)

Consider the homogeneous-consumer logit model:

$$u_{ijt} = \beta_0 + \beta_1 \cdot \text{sugar}_{jt} + \alpha \cdot p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

where:

- u_{ijt}: Utility of consumer i for product j in market t
- sugar_{jt}: Sugar content of product j in market t
- p_{jt}: Price of product j in market t
- ξ_{jt}: Unobserved product quality
- ϵ_{ijt}: i.i.d. Type 1 Extreme Value error

The utility of the outside option (not purchasing any cereal) is normalized to $u_{i0t} = 0 + \epsilon_{i0t}$.

a. **(5 points)** Before running any regressions, what sign do you expect α (the price coefficient) to have? Explain your reasoning.

b. **(5 points)** Using the Berry inversion, write down the linear regression equation that you would estimate. Clearly show the dependent variable on the left-hand side.

c. **(5 points)** Estimate the model using OLS. Report the coefficients $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\alpha}$. (No need to report standard errors.)

d. **(10 points)** Estimate the model using 2SLS, instrumenting for price using the provided instruments (*demand_ instruments0, ..., demand_ instruments19*). Report the coefficients. (No need to report standard errors.)

e. **(10 points)** Compare $\hat{\alpha}$ from 2SLS to your OLS estimate. Which is more negative? Briefly explain why they differ. (2 sentences max.)

f. **(10 points)** Using your 2SLS estimates, compute the own-price elasticity for each product in market 'C01Q1'. Create a scatterplot with prices on the x-axis and own-price elasticities on the y-axis. What pattern do you observe? Is this a feature or a bug of the logit model?

Question 2: Logit with Demographic Interactions (50 points)

Now consider a model that allows preferences to vary with observed consumer demographics:

$$u_{ijt} = \beta_0 + (\beta_1 + \beta_{1,inc} \cdot \text{income}_i) \cdot \text{sugar}_{jt} + (\alpha_0 + \alpha_{inc} \cdot \text{income}_i) \cdot p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

This model has five parameters:

- **Linear parameters:** β_0 (constant), β_1 (sugar), α_0 (price)
- **Demographic interactions:** $\beta_{1,inc}$ (income $\times$ sugar), α_{inc} (income $\times$ price)

Note: This model does *not* include random coefficients—all heterogeneity comes from observed demographics. Again, the utility of the outside option (not purchasing any cereal) is normalized to $u_{i0t} = 0 + \epsilon_{i0t}$.

a. **(10 points)** Explain one advantage that this model has compared to the simple logit model in Question 1.

b. **(10 points)** In the context of this model, what is δ_{jt}? Write out the expression for δ_{jt}.

c. **(20 points)** Estimate this model using `pyblp`. Report all five parameter estimates. (No need to report standard errors.)

d. **(10 points)** Interpret the parameter α_{inc}. (2 sentences max.)

e. **(Bonus: 10 points)** Using your estimates, compute own-price elasticities for each product in market 'C01Q1'. Create a scatterplot as in Question 1(f). How does the relationship between prices and elasticities differ from the basic logit? Explain why this happens.

Submission Checklist

- ☐ PDF with answers to all questions
- ☐ Python code (`.py` or `.ipynb`)
- ☐ All figures clearly labeled
- ☐ Both group members' names (if applicable)