ECN 594: Homework 1

Demand Estimation

Due: See Canvas

You may work in groups of up to 2 people.

Instructions

- 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.

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 (30 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

a. **(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.

b. **(5 points)** Estimate the model using OLS. Report the coefficients $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\alpha}$.

c. **(5 points)** Why might OLS give biased estimates of α? What is the likely direction of the bias? Explain briefly.

d. **(10 points)** Estimate the model using 2SLS, instrumenting for price using the provided instruments (*demand_instruments0, ..., demand_instruments19*). Report the coefficients. Compare $\hat{\alpha}$ to your OLS estimate and comment on the direction of bias.

e. **(5 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?

Hint for part (e): Recall that for the logit model, own-price elasticity is:

$$\eta_{jj} = \alpha \cdot p_j \cdot (1 - s_j)$$

Question 2: Logit with Demographic Interactions (35 points)

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

$$u_{ijt} = \beta_0 + \beta_{0,inc} \cdot \text{income}_i + \beta_1 \cdot \text{sugar}_{jt} + (\alpha_0 + \alpha_{inc} \cdot \text{income}_i) \cdot p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

This model has five parameters:

- **Linear parameters:** β_0 (constant), α_0 (mean price coefficient)
- **Demographic interaction parameters:** $\beta_{0,inc}$ (income effect on baseline utility), α_{inc} (income effect on price sensitivity)
- **Linear parameter:** β_1 (sugar coefficient)

Note: This model does *not* include random coefficients—all heterogeneity comes from observed demographics.

a. **(5 points)** What is the economic interpretation of α_{inc}? What sign would you expect it to have and why?

b. **(15 points)** Estimate this model using `pyblp`. Report all five parameter estimates.

Hint: In `pyblp`, demographic interactions are specified through the "agent formulation." The `agent_data.csv` file contains 20 consumer draws per market. You do not need random coefficients for this problem—set $\Sigma = 0$ and only estimate Π (the demographic interaction matrix). See the pyblp documentation on micro moments for guidance.

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

d. **(5 points)** The IIA property is a limitation of the logit model. Does adding demographic interactions help address IIA? Explain.

Question 3: Consumer Surplus (35 points)

In this question you will compute consumer surplus using your estimated demand system.

a. **(10 points)** Write down the formula for expected consumer surplus per consumer in the logit model. Express it in terms of the "inclusive value" (log-sum).

Hint: For the logit model with utility $V_{jt} = \delta_{jt}$ (mean utility) and outside option normalized to $V_{0t} = 0$:

$$E[CS_t] = \frac{1}{|\alpha|} \ln \left(1 + \sum_{j=1}^{J_t} \exp(\delta_{jt}) \right)$$

b. **(15 points)** Using your 2SLS estimates from Question 1, compute the expected consumer surplus per consumer in market 'C01Q1'.

   Now suppose product 'F1B04' (one of the cereals) is removed from the market. Compute the new consumer surplus. What is the change in consumer surplus from removing this product?

   Note: Assume consumers can still choose the outside option and all other products remain available at the same prices.

c. **(10 points)** The change in consumer surplus from removing a product depends on which product is removed. For market 'C01Q1', identify:

   - The product whose removal causes the *largest* decrease in consumer surplus
   - The product whose removal causes the *smallest* decrease in consumer surplus

   What characteristics do these products have (price, sugar, market share)? Does this make intuitive sense?

Submission Checklist

☐ PDF with answers to all questions

☐ Python code (`.py` or `.ipynb`)

☐ All figures clearly labeled

☐ Both group members' names (if applicable)