Problem Set #1

The goal of this problem set is to learn how to estimate demand and simulate counterfactual scenarios with it. Even though there are some "off-the-shelf" codes available online, you will need to code from scratch. This will be helpful to understand each piece that goes into it the process. You can use python, R, Matlab, or any other coding language that fits you. I will provide solutions coded in Matlab after the due date deadline.

The delivery should be a self-contained pdf with both questions and answers written down in LaTeX. You should also submit a zip file with the codes and the data inside. Make sure to use relative paths so that your code will run completely and at once from any computer. Inside the zip file, you should include a `README.txt` file with clear instructions on how to run the code. The solution needs to be submitted through the portal in bCourse. No late submissions will be accepted. Start early. Don't leave the problem set for the last week.

Through the problem set, you will use different models to estimate the demand for yogurt using the data in `data_yoghurt.csv`. In the data, you will find information about 200 different cities over 36 months. We will define a market as each city-month combination. In each market, there are between 1 and 5 different products provided in Table 1 below.

Table 1: Product characteristics

Product id	Product name	Package size	Calories	Added sugar	Protein
1	Yoplait	170 g	150 kcal	13 g	6 g
2	Chobani	150 g	80 kcal	0 g	14 g
3	Dannon	150 g	100 kcal	0 g	15 g
4	Stonyfield Farm	170 g	150 kcal	11 g	6 g
5	Activia	113 g	90 kcal	13 g	4 g

The data also has information about products' prices (in dollars per unit), market shares, the distance between the city and each product's closest distribution center (in hundred miles), and the average price of diesel during each month (in dollars per gallon).

Chobani and Dannon just announced they want to merge. The local anti-trust agency hired you to determine the potential welfare losses to consumers that can arise from the merge. To simplify the analysis, we will focus on the welfare effects on city 1 during month 1 (we will still use the whole dataset to estimate the model).

1. Logit Demand Model with exogenous prices (15%)

Let's start by assuming that consumer i's utility from consuming a yogurt product j in city c and month t is given by:

$$u_{ijct} = -\alpha p_{jct} + \beta X_j + \xi_{jct} + \epsilon_{ijct} \tag{1}$$

where $\epsilon_{ijct} \sim^{iid} EVI$ and X_j is a vector that contains information about the total grams of each product, and the caloric, sugar, and protein **concentration per gram** of each product. For each market, normalize the utility of the outside option so that $u_{i0ct} = \epsilon_{i0ct}$.

1. Normalize the number of calories, added sugar, and proteins of each product by their package size.[1] Report in a table the price, market share, package size, calories per gram of yogurt, grams of added sugar per gram of yogurt, and proteins per gram of yogurt in city 1 during month 1. What are the two most expensive products? Which ones have the largest market shares?

2. Exploiting the fact that ϵ_{ijct} follows an extreme value type I distribution, write down the formula for a product's market share. Use it to invert the demand equation and derive the demand system in which each equation only depends on a single product's set of characteristics.

3. Assume that $\mathbb{E}[\xi_{jct}|p_{jct}, X_{jct}] = 0$ and estimate the demand system via OLS. Report all coefficients and their EHW robust standard errors.

4. Using the estimated parameters, calculate and report the table and own- and cross-price elasticities of the five products and the outside option in city 1 and month 1. You don't need to report confidence intervals here.

5. Calculate and report the table of diversion ratios. How does the Independence of Irrelevant Alternatives assumption affect these diversion ratios? Do you think these estimates are reasonable?

6. Write down the firm's problem and derive the first-order conditions. Assume that firms only choose prices and that X_j is given.

7. Use the first order conditions to calculate the marginal cost of each product in city 1 and month 1. Report the costs in a Table. Are all costs greater than zero?

8. Explain why assuming that $\mathbb{E}[\xi_{jct}|p_{jct}, X_{jct}] = 0$ can be problematic when recovering the marginal costs.

[1] The normalization is necessary to construct X_j.

2. Logit Demand Model with three way fixed effects (10%)

A friend suggests making use of the fact that you observe the same product in multiple markets and estimate the following model instead:

$$u_{ijct} = -\alpha p_{jct} + \tau_j + \tau_c + \tau_t + \xi_{jct} + \epsilon_{ijct} \tag{2}$$

where τ_j, τ_c, and τ_t are product, city, and month fixed effects.

9. Explain how adding fixed effects can help to solve the problems pointed out in question 8.

10. Estimate the model with fixed effects, report the coefficients, and the own- and cross-price elasticities and diversion ratios for city 1 and month 1. How do they differ from the previous estimates? Why?

11. Report the new estimated marginal costs in city 1 and month 1. Are all the costs greater than zero? Why adding fixed effects might not be enough?

3. Logit Demand Model using cost shifters (15%)

The same friend suggests using differential changes in transportation costs as an instrument for prices. They suggest using the distance between each city and each product's closest distribution center interacted with diesel prices as an instrument.

12. You have a debate on whether to estimate the model with X_j or the model with the three-way fixed effects. Discuss whether estimating the two models is equivalent or not. If not, why? Which one is more likely to provide less biased own-price elasticities? Why?

13. Estimate the model with fixed effects using the newly constructed cost-shifter instrument. Report the coefficients, and the own- and cross-price elasticities and diversion ratios for city 1 and month 1. How do they differ from the previous estimates? Why?

14. Report the new estimated marginal costs in city 1 and month 1. Are all costs greater than zero?

15. Write down an algorithm that allows you to recover the equilibrium prices using the estimated costs and demand parameters. Check that you can recover the prices of all products for city 1 and month 1.

16. Use your algorithm to recover the equilibrium prices in city 1 and month 1 in a counterfactual where Chobani and Dannon merge and marginal costs remain constant. Report the new prices next to the old ones on a table. What happens to prices?

17. Calculate the difference in average consumer welfare in city 1 and month 1 between the scenario with and without the merger. Are consumers better off with the merger?

4. Nested Logit Demand Model (20%)

Your friend is worried that the IIA assumption can be affecting some of the results. Since Chobani and Dannon are the two products with zero sugar content, they think that the cross-price elasticity between the two products might be downward biased under the previous model, which can lead to overestimating the net welfare effects of the merger.

They propose to estimate the same model as in Section 3, using three-way fixed effects and the distance-diesel instrument, but allowing for a more compelling distribution of the shocks ϵ_{ijct}. She proposes to assume that ϵ_{ijct} follows a GEV distribution that leads to a nested logit with three nests: One nest for products without sugar (Chobani and Dannon), another nest for sugary products (Yoplait, Stonyfield Farm, and Activia), and another nest for the outside option.

18. Discuss whether you agree with your friend's intuition that the cross-price elasticity between the two products might be downward biased under the standard logit model. Why? Discuss whether you agree with your friend's intuition that you might have overestimated of the net welfare effects of the merger. Give economic intuition without using math.

19. Using the formula for the market share of the nested logit model, invert the demand equation and derive the demand system in which each equation only depends on a single product's set of characteristics and conditional market share.

20. Estimate the model via 2SLS using as instrument for p_{jct} the cost shifter introduced in Section 3, and for the conditional share function $\log(s_{jct|g})$ the total number of products in city c, month t and nest g. Report the estimated coefficients and explain why we need an additional instrument. What do we need to assume for the instrument to be valid?

21. Calculate and report the own- and cross-price elasticities and diversion ratios for city 1 and month 1. How do they differ from the previous estimates? Why?

22. Report the new estimated marginal costs in city 1 and month 1. Are all the costs greater than zero?

23. Recover the equilibrium prices in city 1 and month 1 in a counterfactual where Chobani and Dannon merge and marginal costs remain constant. Report the new prices next to the old ones on a table. Include the post-merger prices from Section 3 on a third column. What happens to prices after the merger in this model? How do they compare with the post-merger prices of the standard logit model?

24. Calculate the difference in average consumer welfare in city 1 and month 1 between the scenario with and without the merger and compare it to your result of Section 3.

5. BLP Model (40%)

After taking the Industrial Organization class at UC Berkeley, you think that a more flexible model could use random coefficients instead of nests. You also think that estimating a BLP model is super fun and so you would want to do it even if the model was not better.

To make sure we all have the same results, let's focus on the following model specification: Let consumer i's utility be given by

$$u_{ijct} = -\alpha p_{jct} + \tau_j + \tau_c + \tau_t + \beta_i(X_j^s - \bar{X}^s) + \xi_{jct} + \epsilon_{ijct} \tag{3}$$

where $\epsilon_{ijct} \sim^{iid} EVI$, $\beta_i = \bar{\beta} + \sigma v_i$, X_j^s is the sugar content per gram of product of product j, and $\bar{X}^s = \frac{1}{J}\sum_{j=1}^{5} X_j^s$ is the average sugar content in city 1 and month 1.

25. Split utility into a linear and a non-linear part such that

$$u_{ijct} = \delta_{jct} + \sigma v_i(X_j^s - \bar{X}^s) + \epsilon_{ijct}.$$

Write down a code that calculates market shares

$$s_{jct}(\delta_{jct}, \sigma) = \frac{1}{N}\sum_i s_{ijct}(\delta_{jct}, v_i, \sigma)$$

where $N = 50$, and $v_i = \Phi^{-1}(u_i)$.[2] To construct u_i, build a grid of 50 equally distant points that starts in 0.1 and finishes in 0.9.[3]

26. Write down an algorithm that takes a value of σ and the vector of observed market shares s_{jct} and inverts the demand system to recover

$$\delta_{jct}(\sigma) \quad = \quad -\alpha p_{jct} + \tau_j + \tau_c + \tau_t + \xi_{jct} \tag{4}$$

Report the values of δ_{jct} for the five products in city 1 and month 1 when $\sigma = 0$. How do they compare to $\log(s_{jct}) - \log(s_{0ct})$? Why?

27. Find an analytical expression for $\frac{\partial \delta(\sigma)}{\partial \sigma}$

28. Write a BLP algorithm code that:

    (a) Receives a parameter σ, the vector of individual unobservables v_i, the vector of observed market shares s_{jct}, a matrix of endogenous and exogenous variables, X, that contains the vector of prices p_{jct} and the fixed-effects dummies, a vector of instruments, Z, that contains the instruments for the endogenous variables as well as the exogenous variables (i.e., the fixed-effects dummies), a non-linear variable, $(X_j^s - \bar{X}^s)$, and a weighting matrix $W = (Z'Z)^{-1}$.

[2] $\Phi^{-1}(x)$ is the inverse of the CDF of the normal distribution

[3] In Matlab, you can construct v_i by typing `v = norminv(linspace(0.1,0.9,N));`

(b) Uses σ and the vector of observed market shares s_{jct} to recover $\delta_{jct}(\sigma)$

(c) Estimates Equation (4) via linear IV: $[\alpha, \tau]' = ((X'Z)W(Z'X))^{-1}((X'Z)W(Z'\delta(\sigma)))$

(d) Estimates the residuals $\hat{\xi}_{jct}$ from Equation (4): $\hat{\xi} = \delta(\sigma) - X[\alpha, \tau]'$

(e) Estimates the moments $E[\xi_{jct}Z_{jct}]$

(f) Calculates the objective function $G = KE[\xi_{jct}Z_{jct}]'WE[\xi_{jct}Z_{jct}]$, where $K = 10^6$. Multiplying G by K helps us to avoid numerical issues due to small numbers.

(g) Calculates the derivative of the objective function $\frac{\partial G}{\partial \sigma}$

(h) Returns G and $\frac{\partial G}{\partial \sigma}$

You don't need to report the algorithm in the pdf answer but make sure the codes are organized and legible.

29. Build the set of instruments used in Section 3 (the matrix should include the cost-shifter and the fixed-effects dummies) and calculate $G(\sigma = 0)$ using your algorithm. Repeat for $G(\sigma = 10)$. What do you find? Why? Explain why we might have an identification problem.

30. Repeat the exercise but adding an additional instrument. Use as an additional instrument the sum of the sugar content per gram of each product, $z_{ct} = \sum_{j \in (c,t)} X_j^s$. What are the values of $G(\sigma = 0)$ and $G(\sigma = 10)$ now?

31. Plot $G(\sigma)$ for $\sigma = \{-200, -190, -180, ..., -10, 0, 10, 20, ..., 180, 190, 200\}$. Plot $\frac{\partial G}{\partial \sigma}$ for the same range of values. Does $\frac{\partial G}{\partial \sigma}$ cross the zero when $G(\sigma)$ reaches a minimum? Comment about the shape of $G(\sigma)$.

32. Estimate the full model using Newton's gradient method. Report the parameters together with the EHW standard errors in a table.

33. Calculate and report the own- and cross-price elasticities and diversion ratios for city 1 and month 1. How do they differ from the previous estimates? Why?

34. Report the new estimated marginal costs in city 1 and month 1. Are all costs greater than zero?

35. Recover the equilibrium prices in city 1 and month 1 in a counterfactual where Chobani and Dannon merge and marginal costs remain constant. Report the new prices next to the old ones on a table. Include the post-merger prices from Section 3 and 4 on additional columns. What happens to prices after the merger in this model? How do they compare with the post-merger prices of the previous models?

36. Calculate the difference in average consumer welfare in city 1 and month 1 between the scenario with and without the merger and compare it to your result of Section 3 and 4.

37. Comment on the main takeaways of this assignment.