# Problem Set 1

**Due date: October 6th, 2023**

You received by email a dataset that contain sales and price information for a product category (the data comes from the Dominick's database (http://research.chicagobooth.edu/kilts/marketing-databases/dominicks). The dataset you will work with is a subset of the original dataset that has been cleaned to make it more manageable. The data contain:

- store: A unique identifier for each store
- upc: The UPC of the product
- week: An identifier for the week
- move: The number of units sold
- price: The price per size charged for each unit (for example, $/oz)
- profit: The gross profit margin
- descrip: A description of each UPC
- custcoun: The number of customers that shopped at the store in the week
- Brand: A unique identifier for each brand (the first 5 digits of the UPC)

1. Describe the data. Provide interesting statistics that describe the key variables. I am leaving this question a bit open ended. Obviously you want to provide a "standard" table of means, std deviations, etc., of key variables.

   In addition, you should think of useful ways to describe the data. For example, how much of the price variation is over time? How much is across stores? How much is across products? How much price variation is there in the cross-section for each product? If you find little variation in the cross-section, what drives price variation in the cross-section? (hint: can you find a sensible measure of "promotions"?)

   Remember that the data comes from a single supermarket chain. Is there any evidence that the chain engages in uniform pricing? That is, if you fix a product and a week, is there price variation across stores? Do this for each product or brand. Is the supermarket chain more likely to engage in uniform pricing for some brands than others? You may also want to do a variance decomposition to be able to formally address these questions. The goal is to be creative, yet produce interesting and relevant ways to describe the data.

2. Estimate the following (Logit) model

$$u_{ijt} = \alpha p_{jt} + x_j\beta + \xi_{jt} + \varepsilon_{ijt}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T,$$

   where $\varepsilon_{ijt}$ is distributed i.i.d. extreme value. The notation is as in class, with markets denoted by $t$ and corresponding to a store–week combination. For characteristics use UPC dummy variables.

   Define market shares as the quantity divided by the number of shoppers that shopped at each store for each week (the custcoun variable). Estimate the model using OLS and IV methods.

   In principle, there are several IVs that you can use. First, you can use the wholesale price that, because of how the data was created, has to be computed as wholesale $\text{price}_{jt} = \text{price}_{jt} \times \left(1 - \frac{\text{profit}_{jt}}{100}\right)$. Second, as we will discuss later in the course, you could also use "counts" as IVs. For example, you could use the number of other products produced by the same manufacturer, and the number of products produced by rival manufacturers, as instruments for the price of a specific product. Both for the wholesale price

and for the counts you should explain why these may be valid IVs. In this case, however, you will discover that counts will not work as IVs. Why? Report estimated coefficients and (robust) standard errors.

3. Can you use "Hausman" instruments in this application? What would your answer depend on?

4. Estimate the following (nested) logit model

$$u_{ijt} = \alpha p_{jt} + x_j\beta + \xi_{jt} + \zeta_{ig} + (1-\sigma)\varepsilon_{ijt}, \quad i = 1,\ldots,I, \quad j = 1,\ldots,J, \quad g = 1,\ldots,G, \quad t = 1,\ldots,T,$$

where $\zeta_{ig} + (1-\sigma)\varepsilon_{ijt}$ is distributed extreme value. The notation is as in class. For characteristics use UPC dummy variables.

To define the nest structure, put the outside option in group $g = 0$ and all the inside goods in $g = 1$. As before, define market shares as the quantity divided by the number of shoppers that shopped at each store for each week (the custcoun variable). Then define the within-group market share as the total quantity sold by each product relative to the total quantity sold by products in the nest. Estimate the model both by OLS and IV methods. Report estimated coefficients and (robust) standard errors.

5. Using your estimates for both the OLS and IV specifications (for both the simple logit and the nested logit), report own and cross-price elasticities.

To do this, first compute the elasticities for each of the markets and report, using a plot, the density of own and cross-price elasticities across markets. To make things easy, report elasticities for the logit and nested logit in different plots and separate own and cross-price elasticities into different plots. Then, for each model (logit and nested logit), you should have two plots (own and cross-price elasticities), and each plot should report the estimated density for the OLS and IV specifications.

For the cross-price elasticity, for each market compute the mean cross elasticity for each product with respect to all other products and plot the density over using all data for all markets and products. For the own price elasticity you don't need to compute the within-market mean, just plot all of them. For the nested logit, compute the mean cross price elasticity with respect to the prices of the other products in the same nest.

*What can you say about the estimated elasticities? Do the own-price elasticities "make sense?" What can you say about cross-price elasticities? How do the elasticities compare across models? Do the patterns that arise when you move from the Logit to the Nested Logit model make sense? What happens with the own-price elasticities when moving from the Logit to the Nested Logit specification? What happens to the cross-price elasticities? Is this reasonable? Why?*

6. Estimate the following random coefficients Logit model

$$u_{ijt} = \alpha_i p_{jt} + x_j\beta + \xi_{jt} + \varepsilon_{ijt}, \quad i = 1,\ldots,I, \quad j = 1,\ldots,J, \quad t = 1,\ldots,T,$$

where $\varepsilon_{ijt}$ is distributed i.i.d. extreme value. The notation is as in class. For characteristics use UPC dummy variables. Define market shares as in problem set 1. Let

$$\alpha_i = \alpha + \pi \cdot inc_i + \sigma\nu_i$$

where $\alpha$, $\pi$, and $\sigma$ are parameters to be estimated, $inc_i$ is the income of consumer $i$ and $\nu_i$ is an individual specific attribute. Note that the random coefficient is on price only!

Assume that $\nu_i \sim N(0,1)$ and $inc_i$ is distributed log normal with a mean and standard deviation that varies by store. To generate the demographic draws, use the dataset called "demo_stores.csv." In this dataset, the first column is the store id, the second is log median income, and the third is the standard deviation of income. Income represents the log of median income, so you'll need to compute the exponential of what is reported in the data. It is true that the original dataset has the log of

median income and not the mean, but for this problem set that is good enough and you should treat it as the mean. Note that the third column is the standard deviation of the income distribution so it doesn't need to be scaled back.

For the simulation you should use 20 draws of $inc_i$ and $\nu_i$ per store (that is, keep the same draws across weeks for a given store). This is far less than what you would need in a paper but we need to keep things simple here. Estimate the parameters using GMM (one step only). That is, following the notation introduced in class, find $\hat{\theta}_2$ such that

$$\hat{\theta}_2 \in \arg\min_{\theta_2 \in \Theta_2} \ \xi(\theta_2)' Z W^{-1} Z' \xi(\theta_2).$$

You should start from $\delta_0 = \log(s_j/s_0)$ and with random values for $\pi_0$ and $\sigma_0$. Then search over the parameter space to estimate the nonlinear parameters that minimize the GMM objective function. Finally, once you have the nonlinear parameters, recover the linear parameters using the linear IV regression discussed in class. **You do not need to provide standard errors for this part of the problem set**. Use the wholesale price and two lags of the wholesale price as instruments. **Report the constant $\alpha$, $\pi$, and $\sigma$.**

7. Compute own- and cross-price elasticities for the Random Coefficients Logit models. Report them in plots as you did above. How do the own- and cross-price elasticities of the Random coefficients model compare to those of the IV-Logit model?

8. Compute markups and marginal costs under the single-product market structure, the multi-product market structure, and the full monopoly. In one plot, report the distribution of markups under each market structure.

**NOTES:**

- Problem sets must be delivered by email. Send them to fluco@tamu.edu.
- You must turn in an "easy-to-read" and understand file. Regardless of what software you use to prepare the document you will turn in, make sure that all equations, tables, and figures are easy to read and understand. This means that you should make sure all lines in a plot can be distinguished from each other if printed in black and white and the document is held at arms length from your eyes. Same for the tables.
- The file you will turn in must have the following structure: Answer every question in order, including the corresponding tables, figure, and code within the text **before you start answering the following question**. To report the code in a way that it is easy to read, you could use, for example, the "minted" package in LaTeX.
- You must provide comments for your answers. Tables and numbers by itself do not provide meaning. Discuss your estimated elasticities, descriptive statistics, markups, etc. Are these reasonable? Are these what you expected? For instance, how do they compare with estimates reported in the literature?
- Your code must contain comments so that anybody can understand your procedure. Label the variables in an informative way so anyone who reads your code may understand what the variable represents.
- Working in groups on these numerical problem sets is encouraged. All members of a group should ultimately do the calculations and hand them in individually. If you work with someone else, let me know by adding their names below yours on the first page of your file.
- You must code the estimator from scratch. Do not use any pre-packed routine.