Empirical Project: Estimating Discrete Choice Models with Market-level Data

Cristian Huse

Instructions

- Please submit the report in groups of at most 3 (three) persons;
- Since this is a (take-home) evaluation, discussions are allowed only within a group. If a submission has traces of other groups's work, all submissions will be deemed as equally unacceptable and discarded;
- Please note that both the content and the presentation of your work counts;
- Please submit both pdf and Rmd files of your work;
- The report should have at most 20 pages – if unsure, use font Garamond 12pt, single-spaced, 2cm page margins everywhere. Please submit it by email to cristian.huse@uol.de| by 23.59, 25th March 2024.

Overview

In this project your task is to estimate the demand for ready-to-eat cereal based on market-level data. The data isfrom a single cross-sectionalal market and contains information about 50 products and an "outside good". For each observation, the following information is available:

Variable	**Description**
0	Index
Name	Brand and product identifiers
Avg Trans Price	Average transaction price (may include discounts, coupons etc)
Avg Shelf Price	Average shelf price (the price variable to be used)
Avg Ad Expn	Average advertising expenditure, by product
Mkt Share	Market share of product (%)
Sgmnt	Market segment to which the product belongs
Cals	Caloric content of the product, per portion
Fat	Fat content of the product, per portion
Sugar	Sugar content of the product, per portion

Please note that the average transaction price consists of the shelf price minus coupons (very common in the US), thus not reflecting actual product prices.

The file **Cereal_Data.xls** contains the data. Please refer to the topics "Instrumental Variables: Identification" and especially "Discrete Choice Models of Demand with Aggregate Data" for concepts alluded to below.

Questions

1. Provide a brief, non-technical, overview of the market using descriptive tools (tables, plots) and focusing on a lay audience. Which are the salient features of the market, the main brands, segments, and products?

2. **The basics and OLS estimation.** Load the data, define the share of the outside good, create a variable to store the brand information and create the dependent variable to be used in the logit regressions (mean utility). Then estimate the following six specifications using OLS and including product characteristics Price, Cals, Fat, and Sugar:

(1) Without brand fixed-effects, standard errors assuming IID errors;

(2) Without brand fixed-effects, heteroskedasticity-robust standard errors;

(3) Without brand fixed-effects, standard errors clustered at the brand level;

(4) With brand fixed-effects, standard errors assuming IID errors;

(5) With brand fixed-effects, heteroskedasticity-robust standard errors;

(6) With brand fixed-effects, standard errors clustered at the brand level;

Report your results in a table and interpret them. Do you have a preferred specification?

Hint. Use the **fixest** package.

3. **Instruments and IV estimation.** Construct instruments to account for price endogeneity and use them to perform IV estimation. Report estimates on a table with specifications differing according to the instruments used. Which instruments have you constructed, and why? Can one claim that the more the merrier when it comes to instruments in this case?

Note. Use the OLS estimates without and with brand fixed-effects with your preferred standard errors. Instrument each of them with the instruments you constructed and report a table with 6-8 specifications in total.

4. **More instruments and IV estimation.** In case you didn't get a satisfactory specification, can you combine instruments to get better results? If so, how? Report a table with your estimates.

5. **Thinking out of the matrix.** In case you didn't get a satisfactory specification up to now, can you suggest any improvements?

Hint: One improvement comes from a new variable. Another improvement comes from tackling the IIA property of logit models.

As a guidance, When reporting, please include again the following specifications previously reported: (i) your preferred OLS logit specifications with and without brand FEs; (ii) your preferred IV logit specification with and without brand FEs which include all instruments.

6. **Wrapping-up**

(a) Which specification do you suggest, and why? Explain your reasoning in a way that also a layperson (e.g, a judge who never ran a regression) could understand.

(b) Compute the elasticity matrix of your specification. Plot own price elasticities against prices. Plot markups against prices. How to evaluate the specification?