# Empirical Project: Estimating the Demand for Cereals

Cristian Huse

## Instructions

Please work in groups of either two or three persons. Please specify the contribution of each group member in your submission.

The full report should have at most 20 pages – if unsure, use font Garamond 12pt, single-spaced, 2cm page margins everywhere. Please submit it by 23.59, 30th March 2025.

## Overview

In this project you are tasked to estimate the demand for cereal using the a version of the Nevo (2000) data.

The data contains information about breakfast cereals across several geographic markets and time periods. Each row is a product-market combination and each market has the same choice set of breakfast cereals, although with different prices and quantities. The data is in `products.csv`. See the Appendix for a description of the data.

## Questions

### 1. Describe the data

Using `products.csv`, compute summary statistics for different the different variables.

### 2. Compute market shares

To transform observed quantities $q_{jt}$ into market shares $s_{jt} = q_{jt}/M_t$, we first need to define a market size $M_t$. We'll assume that the potential number of servings sold in a market is the city's total population multiplied by 90 days in the quarter. Create a new variable `market_size` equal to `city_population` times `90`. Note that this assumption is somewhat reasonable but also somewhat arbitrary.

Next, compute a new column `market_share` equal to `servings_sold` divided by `market_size`. This gives our market shares $s_{jt}$. We'll also need the outside share $s_{0t} = 1 - \sum_{j \in J_t} s_{jt}$. Create a new column `outside_share` equal to this expression.

## 3. Estimate the pure logit model with OLS

Recall the pure logit estimating equation: $\log(s_{jt}/s_{0t}) = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$. First, create a new column `logit_delta` equal to the left-hand side of this expression.

Then, run an OLS regression of `logit_delta` on a constant, `mushy`, and `price_per_serving`.

(a) Interpret your estimates.

(b) How can you re-express your estimate on `mushy` in terms of how much consumers are willing to pay for `mushy`, using your estimated price coefficient?

(c) Discuss which type of standard errors you should use. Compare their estimates.

## 4. Add market and product fixed effects

Since we expect price $p_{jt}$ to be correlated with unobserved product quality $\xi_{jt}$, we should be worried that our estimated $\hat{\alpha}$ on price is biased. Since we have multiple observations per market and product, and prices vary both across and within markets, it is feasible for us to add both **market and product fixed-effects**.

If $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$ and most of the correlation between $p_{jt}$ and $\xi_{jt}$ is due to correlation between $p_{jt}$ and either $\xi_j$ (product fixed effects) or $\xi_t$ (market fixed effects), then explicitly accounting for these fixed effects during estimation should help reduce the bias of our $\hat{\alpha}$.

Estimate a specification incorporating these fixed-effects and report a table comparing it with the specification without fixed-effects. Interpret any changes.

## 5. Add an instrument for price

Adding market and product fixed effects can be helpful, but since unobserved quality typically varies by both product **and** market, we really want to instrument for prices. The data comes with a column `price_instrument` that we should interpret as a valid instrument for price that satisfies the needed exclusion restriction. It could be a cost-shifter, a valid Hausman instrument, or similar.

(a) Run a first-stage regression to make sure that it's a relevant instrument for price. Does `price_instrument` seem like a relevant instrument for `prices`?

(b) Run your IV regression using `price_instrument` as an instrument for `price_per_serving`. How does your estimate of $\hat{\alpha}$ change?

## 6. Promotion analysis

(a) With the estimates obtained above, compute the own-price elasticities of demand and associated price-cost margins. Discuss their credibility.

Note: Use $\eta_{jj} = -\alpha p_j(1 - s_j)$, as per slides.

Next, your task is to simulate what would happen if we halved an important product's price? To do so, select the most recent quarter in the first city: `C01Q2`.

Create a new dataframe called `counterfactual_data` for just that market and inspect the data. We'll pretend that we're firm one, and deciding whether we want to cut the price of our brand fourth's product `F1B04`.

In your new dataframe with just data from `C01Q2`, create a `new_prices` column that is the same as `prices` but with the price of `F1B04` cut in half.

(b) Compute the percent change in sales of each product in the market. What can you say about the price elasticity of demand of the products?

(c) From firm one's perspective, do the estimates of product cannibalization make sense? That is, do the signs and magnitudes on the demand responses to the price change make sense?

### 7. Nested logit model

(a) Which of the variables is the most suitable to be defined as a nest?

(b) How would you construct the nest shares? Is it exogenous or endogenous?

(c) What would you use if you were to instrument the nest share?

## References

Nevo, A. (2000). "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics & Management Strategy*, 9(4), 513-548.

## Appendix

The columns in the data are as follows.

| Column | Data Type | Description |
|---|---|---|
| `market` | String | The city-quarter pair that defines markets $t$. |
| `product` | String | The firm-brand pair that defines products $j$. There are 5 firms which produce between 1 and 9 brands of cereal. |
| `mushy` | Binary | A dummy product characteristic equal to one if the product gets soggy in milk. |
| `servings_sold` | Float | Total quantity $q_{jt}$ of servings of the product sold in a market, which will be used to compute market shares. |
| `city_population` | Float | Total population of the city, which will be used to define a market size. |
| `price_per_serving` | Float | The product's price $p_{jt}$ used in these exercises. |
| `price_instrument` | Float | An instrument to handle price endogeneity. |

Note that `market` and `product` convey information which you can use once you split the corresponding strings.