

# 1 Organizing Product ids Into Groups

We use a hierarchical structure to organize product codes into groups and categories. One way to do that would be to follow Global Product Classification (GPC): Developed by GS1, the GPC offers a global standard for categorizing products that is more detailed than UPC. It includes definitions for categories and subcategories, making it useful for managing product listings across different regions and languages.

This gives us:

- segment
- family
- Class
- brick

An example in the attached csv is the chocolate bana muffin, which has:

- segment: Food/Beverage/Tobacco
- family: Baked Goods
- Class: Muffins
- brick: chocolate bana muffin

Details on the mapping may be provided here <https://www.gs1.org/>

## 2 Price indeces by Group

I suggest that we aggregate different products hierarchically, starting from the most granualr hierarchy and then aggregating up.

Throughout we use the following notation:

- $i$  denotes the product item, e.g. the most granular product code that we can geometry
- $b(i)$  denotes the brick id for product id  $i$ . I assume a unique mapping from  $i$  to  $b(i)$ .
- $c(i)$  denotes the class id for product id  $i$ . I assume a unique mapping from the brick id  $b(i)$  to the class id  $c(i)$ .
- $f(i)$  denotes the family id for product id  $i$ . I assume a unique mapping from the class id  $c(i)$  to the family id  $f(i)$
- $s(i)$  denotes the segment id for product id  $i$ . I assume a unique mapping from the family id  $f(i)$  to the segment id  $s(i)$

- $t$  denotes time
- $j$  denotes the household (school district) in our application

More notation:

- $Q_i$  denotes the number ordered units. A unit may contain multiple items. For example, a unit may be a box containing 20 ramen noodle packages.
- $Size_i$  denotes the number of items per unit. For instance 20 in the case of ramen noodles
- $Total_i = Q_i \times Size_i$ : denotes the total number of items ordered. If the school orders  $Q=10$  units of ramen noodles, with  $size=20$  items per unit, then a total number of 200 items will be ordered
- $P_{ijt}^Q$  denotes the price per unit (e.g. the price per 20 ramen noodle items). Our data are collected as prices per unit.
- $P_{ijt}^{size}$  denotes the price per serving size, which is the price per item. This is our object of interest

## 2.1 Brick/Class

I would start by assuming that all items within the same brick and class are identical. That means that for every household id (school district) and year, we can simply construct the average price per serving (weighted by the quantity).

For example, we have

$$P_{bjt}^{size} = \frac{\sum_{i \in b} P_{ijt}^Q \times Q_{ijt} \times S_{ijt}}{\sum_{i \in b} Q_{ijt} \times S_{ijt}} \quad (1)$$

$$P_{cjt}^{size} = \frac{\sum_{i \in c} P_{ijt}^Q \times Q_{ijt} \times S_{ijt}}{\sum_{i \in c} Q_{ijt} \times S_{ijt}} \quad (2)$$

We also define base quantities in the first year of data for each school districts:

$$Total_{bj} = \sum_{i \in b, t=t0(j)} Q_{ijt} \times S_{ijt} \quad (3)$$

$$Total_{cj} = \sum_{i \in c, t=t0(j)} Q_{ijt} \times S_{ijt} \quad (4)$$

where  $t0j$  denotes the first year of data for school district  $j$ .

**Goal:** The goal is to define the group (brick/class) sufficiently broadly such that we can construct average prices for every  $j$  and  $t$ .

## 2.2 Family/Segment

How do we aggregate prices to the family level?

### 2.2.1 Intertemporal Analysis

This analysis compares prices within a school district  $j$  over time.

$$P_{fjt}^{size} = \frac{\sum_{b \in f} P_{bjt}^{size} \times Total_{bj}}{\sum_{b \in f} Total_{bj}} \quad (5)$$

$$P_{sjt}^{size} = \frac{\sum_{b \in s} P_{bjt}^{size} \times Total_{bj}}{\sum_{b \in s} Total_{bj}} \quad (6)$$

These price indeces hold quantities fixed over time but allow prices in brick to vary.

### 2.2.2 Cross-Sectional Analysis

This analysis compares prices across school district  $j$  within a given year.

To this end, we anchor quantities around an overall average in a given year. Below, we sum over  $i$  and over  $j$ :

$$Total_{bt}^{cross} = \sum_{i \in b, j} Q_{ijt} \times S_{ijt} \quad (7)$$

$$Total_{ct}^{cross} = \sum_{i \in c, j} Q_{ijt} \times S_{ijt} \quad (8)$$

We then have

$$P_{fjt}^{size, cross} = \frac{\sum_{b \in f} P_{bjt}^{size} \times Total_{bt}^{cross}}{\sum_{b \in f} Total_{bt}^{cross}} \quad (9)$$

$$P_{sjt}^{size, cross} = \frac{\sum_{b \in s} P_{bjt}^{size} \times Total_{bt}^{cross}}{\sum_{b \in s} Total_{bt}^{cross}} \quad (10)$$