Elementary aggregation: A not so elementary story!

Claude Lamboray

16th meeting of the Ottawa Group
May 7 – 10, 2019
Rio de Janeiro, Brazil
Background

Traditional CPI compilation paradigm:
- *Prices are aggregated without weights at the level of the elementary aggregate.*
- *These elementary price indices are then aggregated to the higher-levels using expenditure weights.*

What does “elementary aggregation” mean in the context of scanner data?
What is an elementary aggregate?

An elementary aggregate is
- the smallest aggregate for which expenditure data are used for CPI purposes.
- the values of the elementary aggregates are used to weight the price indices for elementary aggregates to obtain higher-level indices.
- the range of goods and services covered by an elementary aggregate should be relatively narrow.
- Elementary aggregates also serve as strata for the sampling of prices.

What is an item?

• *In practice, the item corresponds to*

  • an *individual product*, specified by its characteristics, and for which *a price is collected in an outlet* at a given *time*

  or

  • a set of transactions which refer to *one or more individual products*, *from one or more places of purchase*, over a *period of time*, and for which *an average price can be computed*. 
### Aggregation structures (1)

<table>
<thead>
<tr>
<th>Level</th>
<th>Aggregation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI product category</td>
<td>Laspeyres-type</td>
</tr>
<tr>
<td>CPI product category by outlet-type and/or by region = EA</td>
<td>Laspeyres-type</td>
</tr>
<tr>
<td>Items</td>
<td>Jevons</td>
</tr>
<tr>
<td>Sub-items</td>
<td>Average price</td>
</tr>
</tbody>
</table>

*Other fixed weights may be used to aggregate the prices of the sampled items.*
### Aggregation structures (2)

<table>
<thead>
<tr>
<th>Index</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Jevons</strong></td>
<td>$I = \left(\frac{p_1^t}{p_1^0} \cdot \frac{p_2^t}{p_2^0} \cdot \frac{p_3^t}{p_3^0} \cdot \frac{p_4^t}{p_4^0} \cdot \frac{p_5^t}{p_5^0}\right)^{\frac{1}{5}}$</td>
</tr>
<tr>
<td><strong>Jevons + Geo. Lasp.</strong></td>
<td>$I = \left(\frac{p_1^t}{p_1^0} \cdot \frac{p_2^t}{p_2^0} \cdot \frac{p_3^t}{p_3^0}\right)^{\frac{1}{3}}w_1 \cdot \left(\frac{p_4^t}{p_4^0} \cdot \frac{p_5^t}{p_5^0}\right)^{\frac{1}{2}}w_2$</td>
</tr>
<tr>
<td></td>
<td>$= (I_1)^{w_1} \cdot (I_2)^{w_2}$</td>
</tr>
<tr>
<td><strong>Jevons + Lasp.</strong></td>
<td>$I = w_1 \left(\frac{p_1^t}{p_1^0} \cdot \frac{p_2^t}{p_2^0} \cdot \frac{p_3^t}{p_3^0}\right)^{\frac{1}{3}} + w_2 \cdot \left(\frac{p_4^t}{p_4^0} \cdot \frac{p_5^t}{p_5^0}\right)^{\frac{1}{2}}$</td>
</tr>
<tr>
<td></td>
<td>$= w_1I_1 + w_2I_2$</td>
</tr>
</tbody>
</table>
**Aggregation structures (3)**

<table>
<thead>
<tr>
<th>Level</th>
<th>Aggregation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI product category</td>
<td>Laspeyres-type</td>
</tr>
<tr>
<td>Retailer</td>
<td>Laspeyres-type</td>
</tr>
<tr>
<td>Product sub-category</td>
<td>?</td>
</tr>
<tr>
<td>Items</td>
<td>Multilateral</td>
</tr>
<tr>
<td>Transactions</td>
<td>Unit value</td>
</tr>
</tbody>
</table>

*Detailed strata with possibly fixed weights may or may not be defined below the level of the CPI product category.*
<table>
<thead>
<tr>
<th>Aggregation structures (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Index</strong></td>
</tr>
<tr>
<td>CCDI in <strong>1 step</strong>, up to the category</td>
</tr>
<tr>
<td>CCDI + Geo. Lasp. in <strong>2 steps</strong>, first to the sub-category and then to the category</td>
</tr>
<tr>
<td>CCDI + Lasp. in <strong>2 steps</strong>, first to the sub-category and then to the category</td>
</tr>
<tr>
<td>CCDI + Törnqvist in <strong>2 steps</strong>, first to the sub-category and then to the category</td>
</tr>
</tbody>
</table>
What is the impact of fixed weights?

- Some simplifying assumptions:
  - The set of items is constant over time.
  - Törnqvist instead of CCDI.
What is the impact of fixed weights?

The difference between the 2-step Törnqvist + Geo. Lasp. index and the 1-step Törnqvist index can be decomposed into three terms.

\[
\text{Covariance}(\ldots, \ldots) + \text{Covariance}(\ldots, \ldots) + \text{Covariance} (\text{Average price of the sub-category}, \\
\text{Deviation of the sub-category fixed weights from the sub-category “true” weights})
\]
Empirical analysis

• *Simulations performed on Dominick’s data set (Mehrhoff (2019)).*

• *Data aggregated across all stores, transformed into monthly data, using Dominick’s item code.*

• 6 categories: *dish detergents, soft drinks crackers, cookies, grooming products, cheese.*

• *For each category, sub-categories are constructed using the pre-defined Dominick's Commodity Code.*
Index compilation

• Indices compiled for 25 months (April 1995 – April 1997)
• 1-step index up to the category: CCDI
• 2-step index, first sub-category and then category: CCDI-Laspeyres, CCDI-Geo. Laspeyres, CCDI-Törnqvist
• Price reference period: April 1995
• Weight reference period for the fixed sub-category weights: April 1994 - March 1995

Dividing the category dish detergents into three sub-categories according to the commodity code.
Example: Dish detergents
A variant of the CCDI index

1. Within each sub-category k, compile matched bilateral Törnqvist indices $P_{i,j}^k$ between any two periods i and j.

2. Aggregate these bilateral indices to the category level as follows.

$$\tilde{P}_{i,j} = \prod_{k=1}^{K} (P_{i,j}^k)^{0.5*\left(w_{i,j}^k(i) + w_{i,j}^k(j)\right)}$$

$$w_{i,j}^k(i) = \frac{\sum_{\text{Item} \in S_i^k \cap S_j^k} p_{\text{Item}} q_{\text{Item}}}{\Sigma_r \sum_{\text{Item} \in S_i^r \cap S_j^r} p_{\text{Item}} q_{\text{Item}}}$$

$$w_{i,j}^k(j) = \frac{\sum_{\text{Item} \in S_i^l \cap S_j^l} p_{\text{Item}} q_{\text{Item}}}{\Sigma_r \sum_{\text{Item} \in S_i^r \cap S_j^r} p_{\text{Item}} q_{\text{Item}}}$$

3. Use the indices $\tilde{P}_{i,j}$ as building blocks to compile a CCDI index ($\tilde{P}_{i,j}$ satisfies the time reversal test).
Example: Dish detergents
A more “standard” price index

• **Cut-off Jevons + Laspeyres index:**
  1. Within each sub-category, consider items available in all 25 months.
  2. Within each sub-category, select the n=10 items with the largest sales.
  3. For each sub-category, compile a Jevons price index over these items.
  4. Aggregate the sub-category Jevons indices using a Laspeyres-type formula.

Such an index tends to be closest to a CCDI-Laspeyres index.
Example: Dish detergents
Conclusions

With scanner data one needs to:

1. Define the item
2. Aggregate the prices of the items up to an intermediate level
3. Aggregate the intermediate elementary price indices

We compiled various 2-step indices and compared them to a 1-step index which does not take into account an intermediate level.
Conclusions

- How “narrow” should categories be constructed at the first stage of aggregation?

- Focus on the product dimension: what about the outlet dimension?

- “Consistency in aggregation”, multilateral methods and dynamic universe?

- A more standardized way for describing elementary aggregation in a CPI?