A new methodology for scanner data in the Dutch CPI

Meeting of the Group of Experts on Consumer Price Indices
UNECE, Geneva, 2-4 May 2016

Antonio Chessa (CPI)
Outline

• Scanner data in the Dutch CPI
• New methodology:
  1. Product definition
  2. Index method
• Some results
• First experiences in CPI
## Use of scanner data in Dutch CPI (2015)

<table>
<thead>
<tr>
<th>Retailer</th>
<th>Scanner data</th>
<th>Survey data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supermarkets</td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>Do it yourself stores</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Department store</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Drugstores</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Travel agencies*</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Fuel*</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>Mobile phones*</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>78.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>21.1</strong></td>
<td><strong>78.9</strong></td>
</tr>
</tbody>
</table>

In % of Coicop weights 2015

* Electronic transaction data, not specified by GTIN
## Current methods: Properties and issues

<table>
<thead>
<tr>
<th></th>
<th>Supermarkets</th>
<th>Other retailers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data processing</strong></td>
<td>All GTINs, subject to filters (dump prices, ‘low’ turnover)</td>
<td>Samples of GTINs (% turnover)</td>
</tr>
<tr>
<td><strong>Item replacement</strong></td>
<td>Old and new GTINs are not matched; dump prices are removed</td>
<td>Manual replacement in some cases, which is time consuming</td>
</tr>
<tr>
<td><strong>Index method</strong></td>
<td>Monthly chained Jevons ⇒ equal weights in elementary aggregates</td>
<td>Laspeyres type index</td>
</tr>
</tbody>
</table>
New methodology: Aims

• Generic method, applicable to different goods/retailers
• Integral data processing, timely inclusion of new items
• Link outgoing GTINs to follow-up items automatically
• Expenditure weights at product level
• Reduce use of filters (preferably no filters at all)
Product definition

• GTINs represent the most detailed level of differentiation

• Can we say: GTIN = product?
Problem with GTINs: “Relaunches”

GTIN: 36-00521-74076-7
Elvive shampoo 2-in-1 multivitamine
Content: 250 ML
Price week 38, 2011: €3,18
Price week 39, 2011: €2,00

GTIN: 36-00522-00499-8
Elvive shampoo 2-in-1 multivitamine
Content: 250 ML
First sold in week 39, 2011
Price week 39, 2011: €3,98
Impact on price index shampoo

From: Chessa A.G., Comparing scanner data and survey data for measuring price change of drugstore articles.
Linking of GTINs

• Not an issue for “stable” assortments, no relaunches

• Otherwise, there are two linking possibilities:

  1. *Retailers’ product codes*:
     - Stock Keeping Units (SKU’s)
     - Test data set of do-it-yourself stores analysed
     - Promising results, full data will be supplied

  2. *Item attributes*:
     - GTINs are combined that share the same characteristics
     - Which attributes should be selected?
Index method: Main properties

• Price index = Turnover index ÷ Quantity index

• *Quantities sold are weighted*:
  - By keeping product prices fixed
  - Prices from *current* publication year are used
  - Prices from multiple periods are included in product weights ⇒ no price imputations for new items are needed
  - Prices of different periods are deflated by price index
Formulas

• Price index:

\[ P_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t}}{\sum_{i \in G_t} v_i q_{i,t}} \]

\[ = \frac{\bar{p}_t/\bar{p}_0}{\bar{v}_t/\bar{v}_0} \]

“Unit value index”

\[ v_i = \sum_{z \in T} \varphi_{i,z} \frac{p_{i,z}}{P_z}, \quad \varphi_{i,z} = \frac{q_{i,z}}{\sum_{s \in T} q_{i,s}} \]

• Geary-Khamis method in IPC (PPPs)

• However, the results look robust under different forms and weighting schemes for the \( v_i \)
Index calculation and updating problem

• *Iterative scheme:*
  - Set initial values for price indices in each month
  - Recalculate the $v_i$ and price indices until they converge

• *How to proceed from month to month?*
  - Expand time window w.r.t. fixed base month
  - Update the $v_i$ with prices and quantities of the current month
  - Calculate *direct* indices with updated $v_i$ w.r.t. base month

• Price indices equal a transitive “benchmark index” in the course of a year $\Rightarrow$ no chain drift
Real time vs transitive benchmark index

- Benchmark index has yearly fixed product weights
- Real time index is calculated with monthly updated weights
- Data used: Scanner data of a Dutch department store
Impact of different weighting in $v_i$

**Department store**

Feb. 2009 = 100

**Mobile phones**

Dec. 2013 = 100

Base method  Expenditure share weighted Vi  Equal weights  Base method  Expenditure share weighted Vi  Equal weights
First experiences in CPI

• QU-method in production for smartphones (Jan. 2016)

• *Big improvement over previous method:*
  - Tighter product definitions
  - Products reflect user experience (performance, storage)
  - New products timely included
  - Only 30-45 minutes of monthly work (2-3 days for old method)

• Next: department store, do-it-yourself stores, drugstores
Contact details

Antonio Chessa
Statistics Netherlands (CPI)
E-mail: ag.chessa@cbs.nl