

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)



Statistics
Netherlands

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)

Retailer	Scanner data	Survey data
Supermarkets	13.5	
Do it yourself stores	0.5	0.9
Department store	0.7	
Drugstores	0.6	
Travel agencies*	1.7	
Fuel*	3.6	
Mobile phones*	0.5	
Other		78.0
Total	21.1	78.9

in % of Coicop weights 2015

* Electronic transaction data, not specified by GTIN

Current methods: Properties and issues

	Supermarkets	Other retailers
<i>Data processing</i>	All GTINs, subject to filters (dump prices, 'low' turnover)	Samples of GTINs (% turnover)
<i>Item replacement</i>	Old and new GTINs are not matched; dump prices are removed	Manual replacement in some cases, which is time consuming
<i>Index method</i>	Monthly chained Jevons \Rightarrow equal weights in elementary aggregates	Laspeyres type index

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

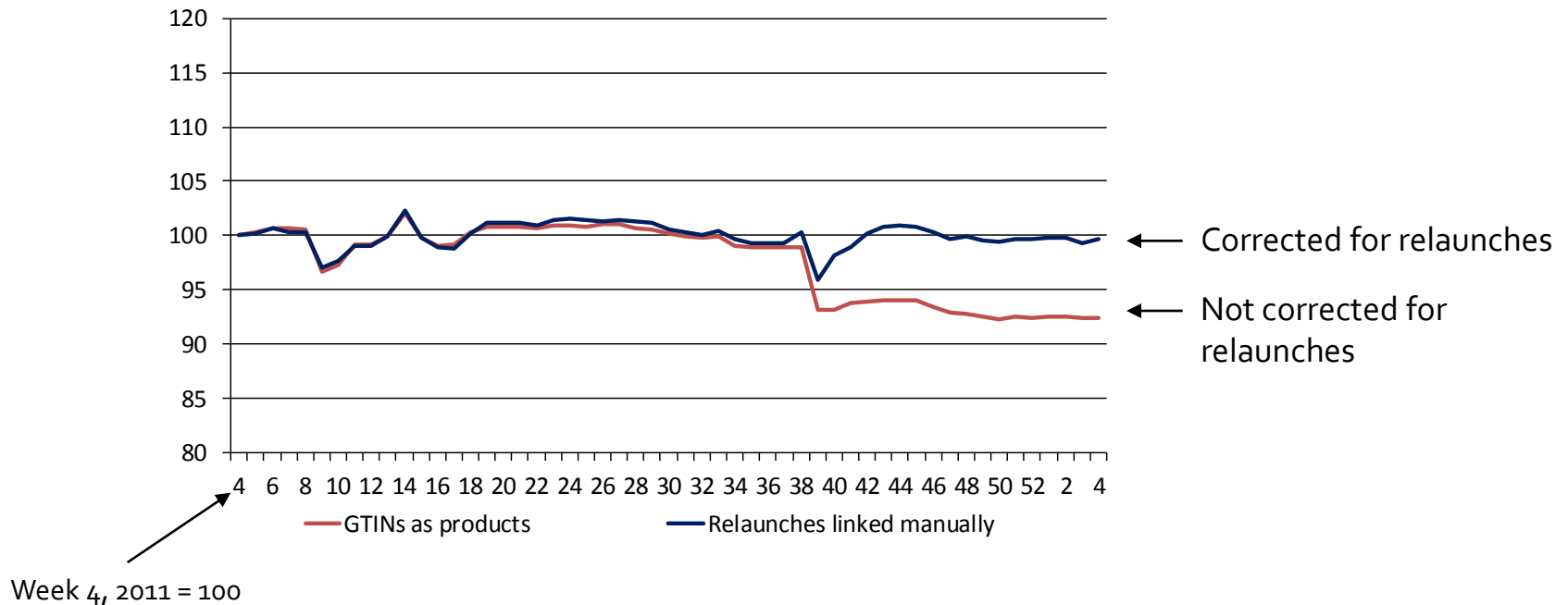
Elvive shampoo 2-in-1 multivitamine

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. Paper presented at the *Workshop on Scanner Data for HICP*, 26-27 September 2013, Lisbon.

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 \div 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 \Rightarrow no price imputations for new items are needed
 - Prices of different periods are deflated by price index

Formulas

- Price index:

$$P_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}} = \frac{\bar{p}_t / \bar{p}_0}{\bar{v}_t / \bar{v}_0}$$

← Unit value index
← "Quality index"

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

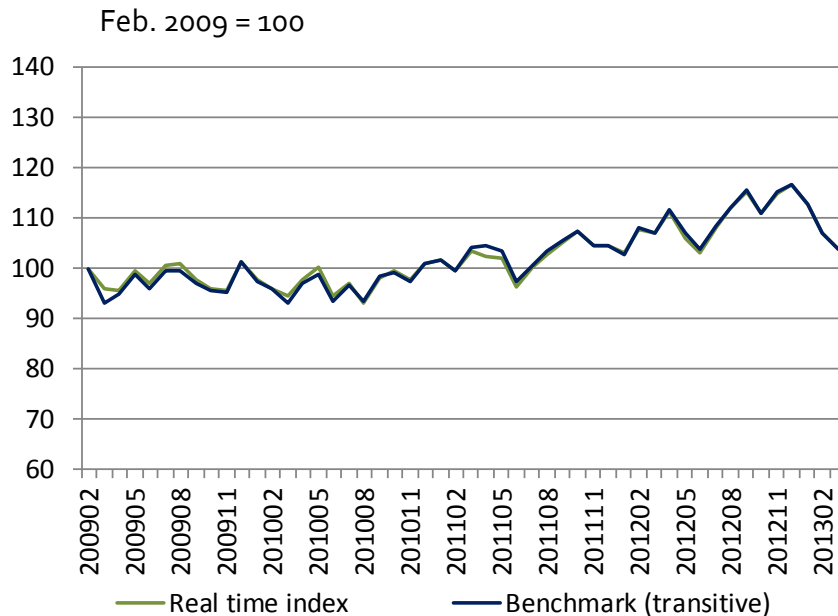
- Geary-Khamis method in IPC (PPPs)
- However, the results look robust under different forms and weighting schemes for the v_i

Index calculation and updating problem

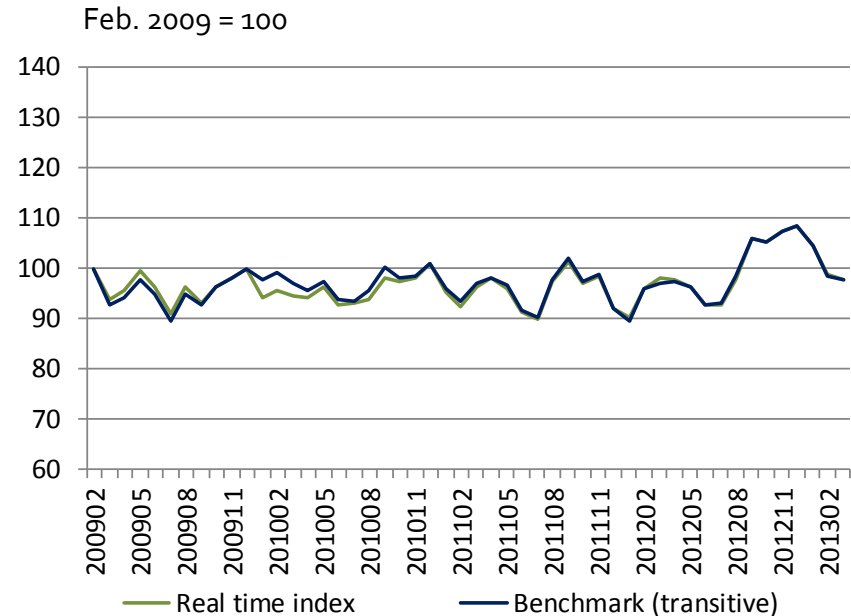
- *Iterative scheme:*
 - Set initial values for price indices in each month
 - Recalculate the v_i and price indices until they converge
- *How to proceed from month to month?*
 - Expand time window w.r.t. fixed base month
 - Update the v_i with prices and quantities of the current month
 - Calculate 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

Men's clothing



Ladies' clothing

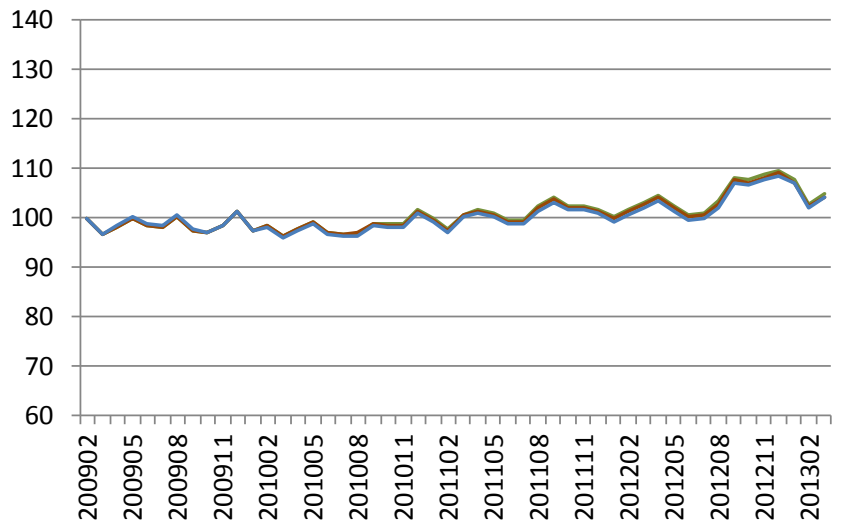


- Benchmark index has yearly fixed product weights
- Real time index is calculated with monthly updated weights
- Data used: Scanner data of a Dutch department store

Impact of different weighting in v_i

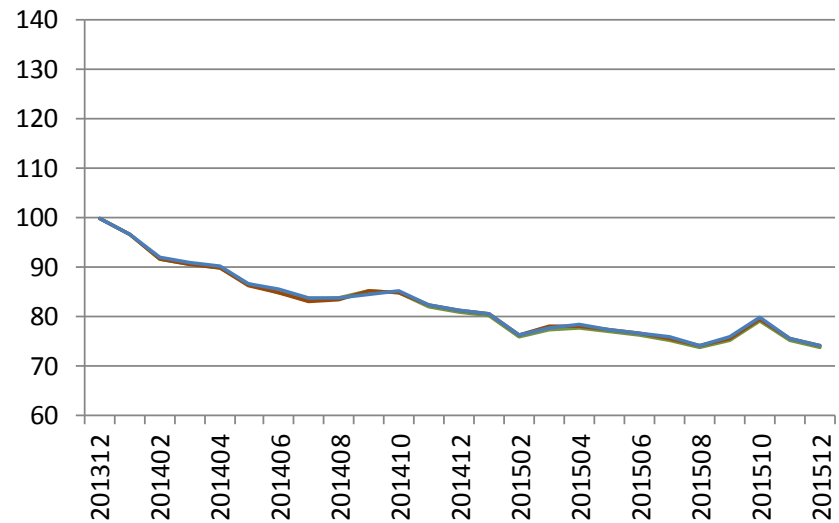
Department store

Feb. 2009 = 100



Mobile phones

Dec. 2013 = 100



— Base method — Expenditure share weighted V_i — Equal weights — Base method — Expenditure share weighted V_i — 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