Towards a new methodology for scanner data in the Dutch CPI

Antonio Chessa, Stefan Boumans and Jan Walschots

Scanner data workshop, Rome, 1-2 Oct. 2015

Statistics Netherlands

Outline

- Historical background
- Motivation and aims of new methodology
- Integration in CPI framework
- The index method
- Results
- Future plans

Scanner data vs Survey data

Scanner data

- Data on all transactions
- Order of 10-100,000 EANs
- Automised price collection
- Turnover
- Quantities sold
- Info about article characteristics in EAN descriptions and/or in separate fields

Issues:

- Returned articles
- Discounts

Survey data

- Sample of price observations
- Several 100s of products
- Prices collected by price collectors
- Prices observed in shops
- Sales are not available
- Consumer specialist sets up article descriptions

Methods for supermarket scanner data

	Version 0	Version 1	Version 2
Developed/used in	Late 1990s	2002-2009	2010-present
Sample	All data	Basket (± 10,000 EANs per retailer)	All EANs that satisfy certain filters
Homogeneous products	EANs	EANs	EANs
Replacements	No	Yes, manually (EANs with large turnover share)	No
Index method	Monthly chained Fisher index	Laspeyres, with yearly fixed weights	Monthly chained Jevons, with equal weights for 'accepted' EANs
Implemented?	No	Yes	Yes

Issues with Version 2

- Filter settings (need to be tested)
- *Relaunches* may lead to downward bias of price index
- Price dump filter is used, but EANs are not matched

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

EAN: 36-00522-00499-8
Elvive shampoo 2-in-1 multivitamine
Content: 250 ML
In week 39 sold for first time
Price week 39: **€ 3,98**

Electronic transaction data in CPI 2015

in % of Coicop weights:

Retailers	Transaction data	Survey data
Supermarkets*	13.5	
Do it yourself stores*	0.5	0.9
Department stores*	0.7	
Drug stores*	0.6	
Travel agencies	1.7	
Fuel	3.6	
Mobile phones	0.5	
Other		78.0
Total	***21.1***	***78.9***

* Scanner data, i.e. transaction data specified by EAN/GTIN

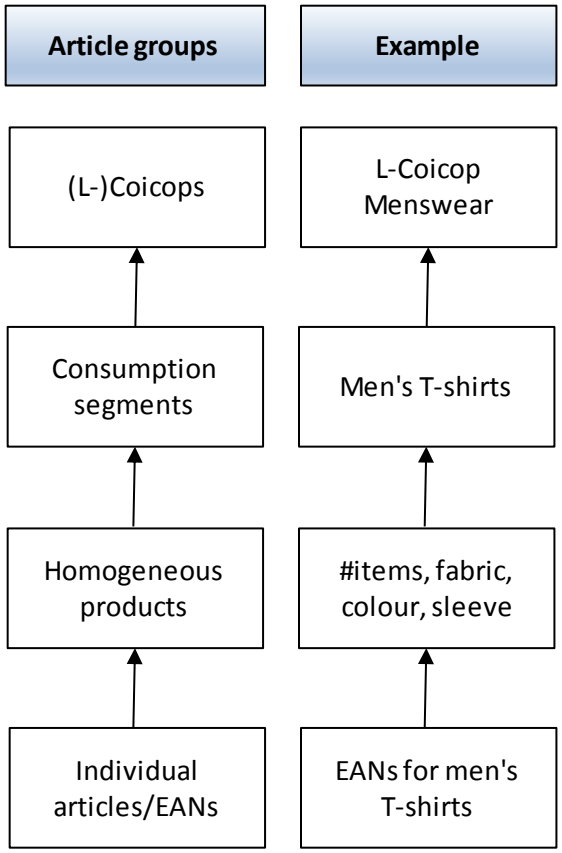
Research objective:

Investigate whether a generic method can be developed that can be applied to scanner data of different retailers/consumer goods

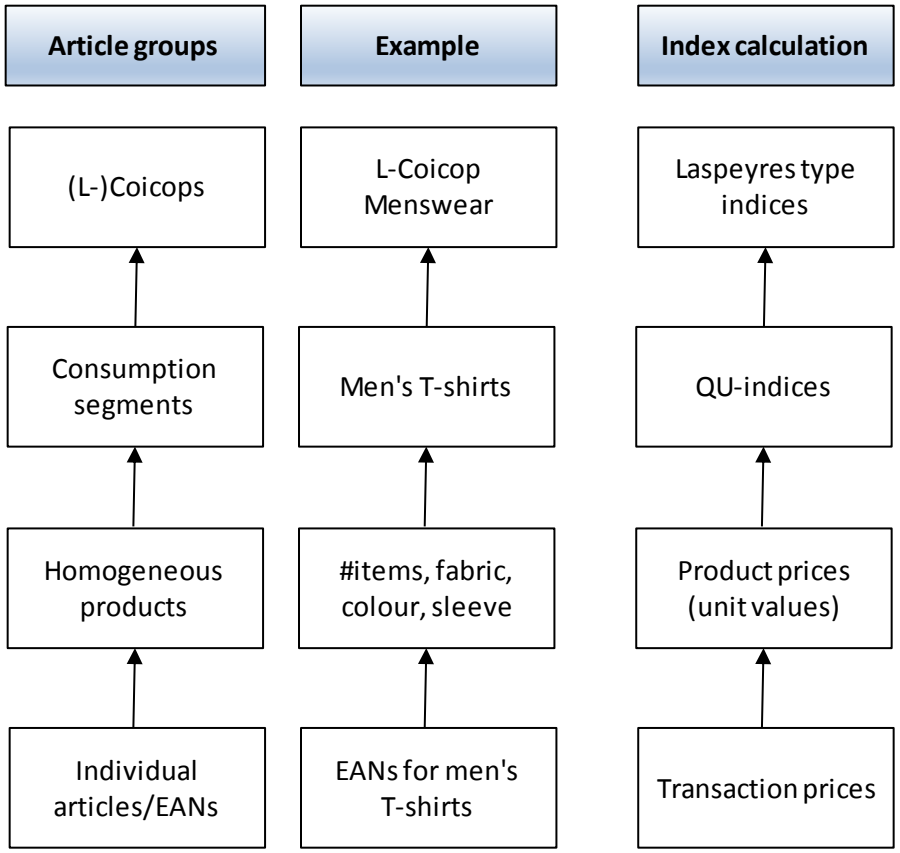
Specific focus:

- Process all articles/EANs
- Timely include new articles
- Reduce use of filters
- Handle *relaunches*

New methodology within CPI



New methodology within CPI



Index method ("QU-method")

- *Definition:*

  Value index ÷ Weighted quantity index

$$P_t = \frac{\sum_{i \in G_t} p_{i,t}\, q_{i,t} / \sum_{i \in G_0} p_{i,0}\, q_{i,0}}{\sum_{i \in G_t} v_i\, q_{i,t} / \sum_{i \in G_0} v_i\, q_{i,0}}$$

0 = base month
t = publication month
G = consumption segment
i = homogeneous product

- *Some special cases:*
  - If $v_i = p_{i,t}$ for all products $i \Rightarrow$ Laspeyres index
  - If $v_i = p_{i,0}$ for all products $i \Rightarrow$ Paasche index
  - All products homogeneous $\Rightarrow$ Unit value index

QU-index is an adjusted unit value 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

← Shift in 'product mix'

Choices about v_i

- Form: $v_i = \sum_{z \in T} \varphi_{i,z} \frac{p_{i,z}}{P_z}, \qquad \varphi_{i,z} = \frac{q_{i,z}}{\sum_{s \in T} q_{i,s}}$

- The v_i are allowed to vary from year to year
- Fixed base month (December of each year)
- Choices based on statistical and sensitivity analyses

For more details, see paper:
Chessa A.G., Towards a generic price index method for scanner data in the Dutch CPI. *Ottawa Group Meeting*, 20-22 May 2015, Urayasu City, Japan.

Computational method

- Iterative method (alternate updates of index and product weights)
- Monthly update of weights with product prices and quantities
- Direct index with updated weights

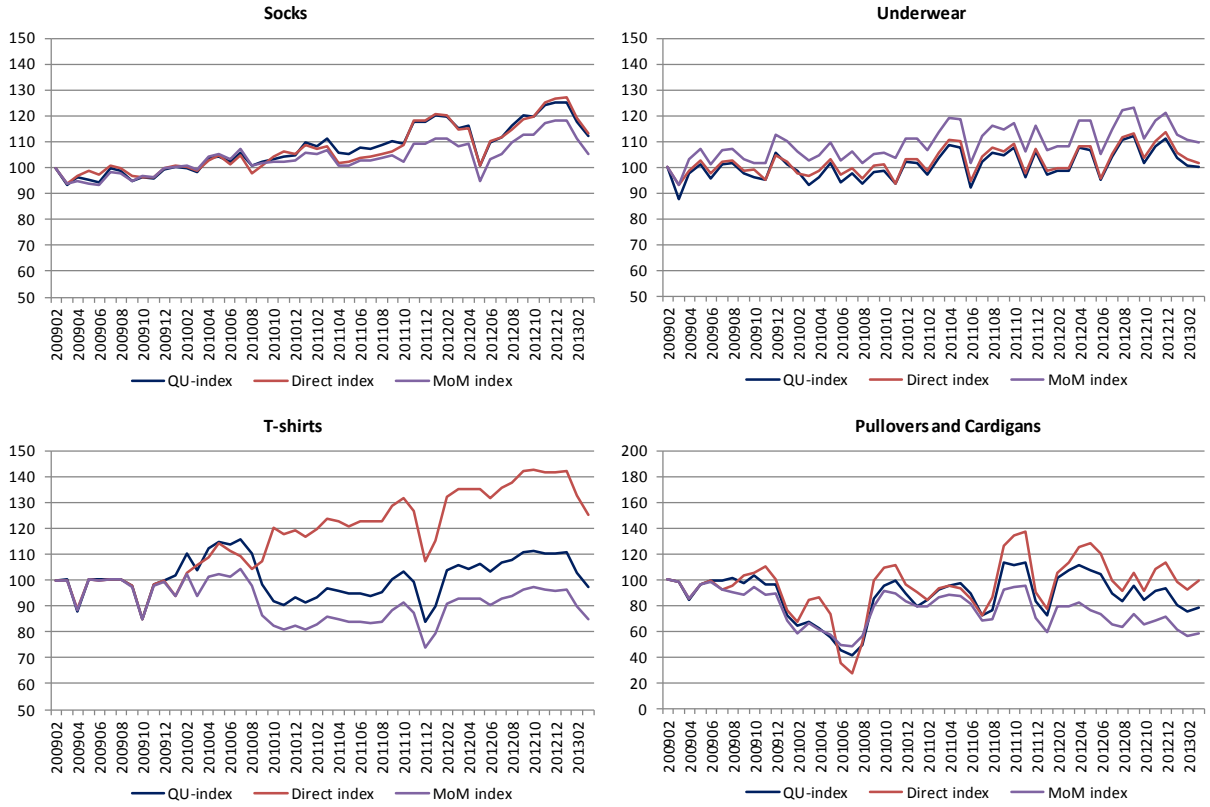
- Index with yearly fixed product weights is used as benchmark (transitive)

Tests

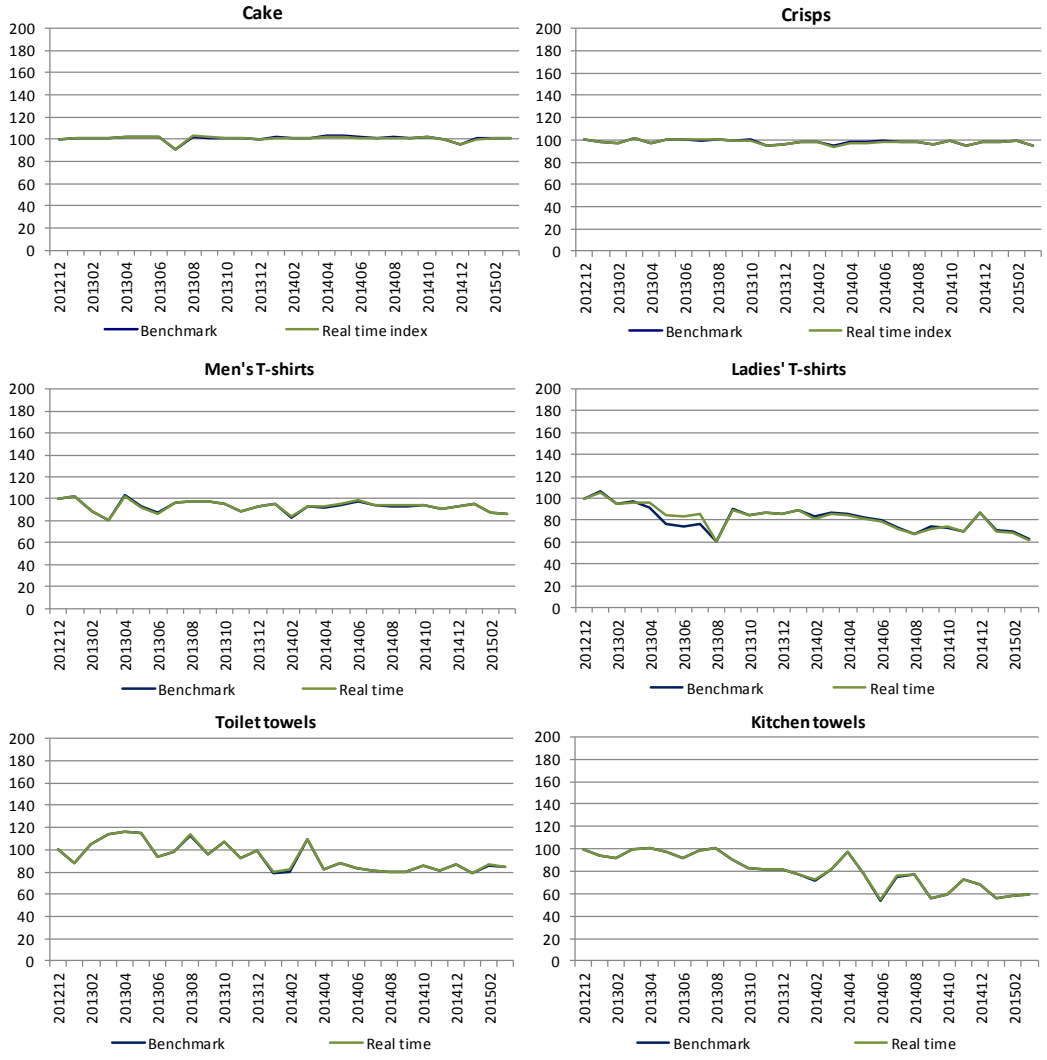
- Department store scanner data and mobile phone data
- Results validated
- Move towards CPI production for both data sets by January 2016

Results: Contribution of new products

Price indices for menswear, department store scanner data (Feb. 2009 = 100)



Test results and benchmarking



Department store scanner data (Dec. 2012 = 100)

Benchmark index has yearly fixed product weights

"Real time index" makes use of monthly updated weights

Short term plans

- *Current state:*
  - Methodology tested for department store and mobile phones
  - Work is in progress for other sectors
- *Future plans and aims:*
  - January 2016: In production for dept store and mobile phones
  - From second half of 2015:
    - Method applied to drugstore scanner data
    - Additional data needed (discounts, maybe also on article characteristics)
    - Preliminary research for DIY-stores (additional data was needed, test data received)
    - Started with supermarkets (small scale research)