Panel on Improving Cost-of-Living Indexes and Consumer Inflation Statistics in the Digital Age

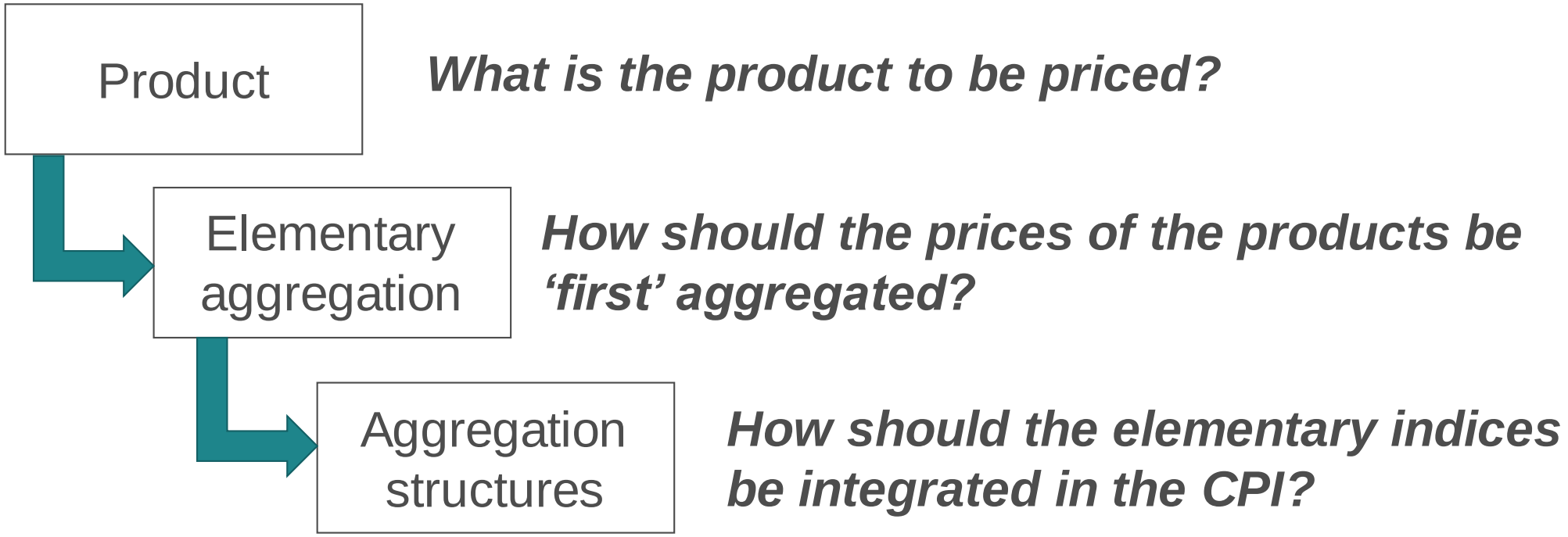
Open Session: Use of high-frequency and other alternative data sources in price measurement

October 7, 2020

Claude Lamboray | Eurostat

claude.lamboray@ec.europa.eu

Outline



European Commission

What is the product to be priced ?

The three main dimensions: time, outlet, product

European Commission

What is the product to be priced? Time

- Unit values for **scanner data** over the reference period (e.g.) month
- Unweighted averaging of price quotes in the case of **web-scraped data** ?
- **Field price collectio**n: prices collected at a certain point in time

Further complications of the time dimension for some products (airfares etc.)?

What is the product to be priced? Outlet

- Is it best to **work at outlet level** or is **aggregation across outlets** (of the same type/chain/region) acceptable?

European Commission

What is the product to be priced? Product

- Use of GTIN/SKU/UPC as a **product identifier**, sometimes only broader product identifier are available in the data sets (leading to some unit value bias)
- ***'The primary obstacle to dealing with transaction data in the CPI has been dealing with product lifecycle effects.'*** *(Big Data in the U.S. Consumer Price Index: Experiences & Plans)*
  - Matching of outgoing and incoming item codes in order to capture relaunches
  - Adjust for package size at this level (for example chocolate bar from 80g to 75g)
  - Dumping filters
  - Combining brand and generic pills (CorpY), homogenous products
- Which type of **discounts** are captured in the standard CPI methodology and which ones are captured in the ADS?

Elementary aggregation

Multilateral methods instead of bilateral methods?

European Commission

Elementary aggregation: Bilateral Methods

- **Unweighted** index formulas: Jevons
  - Use either all the data, or only a sample (weights can be taken into account in the sampling process)
- **Weighted** index formulas: Geometric Laspeyres, Törnqvist,…
  - Frequently chaining is not recommended

European Commission

Elementary Aggregation: Multilateral Methods

- **Multilateral method** as a way to take into account all the data (prices and weights), and cope with a dynamic product universe
- Examples:
  - Belgium: GEKS-T; 25 month rolling window; Half splice
  - Australia: GEKS-T; 9 quarters rolling window; Mean splice
  - UK: GEKS-Jevons; Movement Splice; for web scraped data
- **Hedonic extensions of the multilateral methods**, taking into account prices, weights and product characteristics

Elementary Aggregation: Discussion

How should we decide on a (multilateral) index?

- Some **broad principles** (UK, Australia): Resources, Theoretical properties, Flexibility, Interpretability, Transitivity versus characteristicity

More concretely:

- The **link with bilateral indices** (GEKS-T/CCDI, GEKS-Jevons)
- **Theory**: For example substitution bias of multilateral index formulas
- **Test approach**: to be further investigated
- A lot of **empirical work**! (but what to conclude from it !?)

European Commission

Elementary Aggregation: Discussion

Some challenges:

- For multilateral methods: decisions must be made on technical aspects such as **window length, splicing technique**
- How do the new compilation techniques relate to the usual CPI treatments: **seasonal products**, **replacements and quality adjustment**
- For web scraping: the **absence of weights**

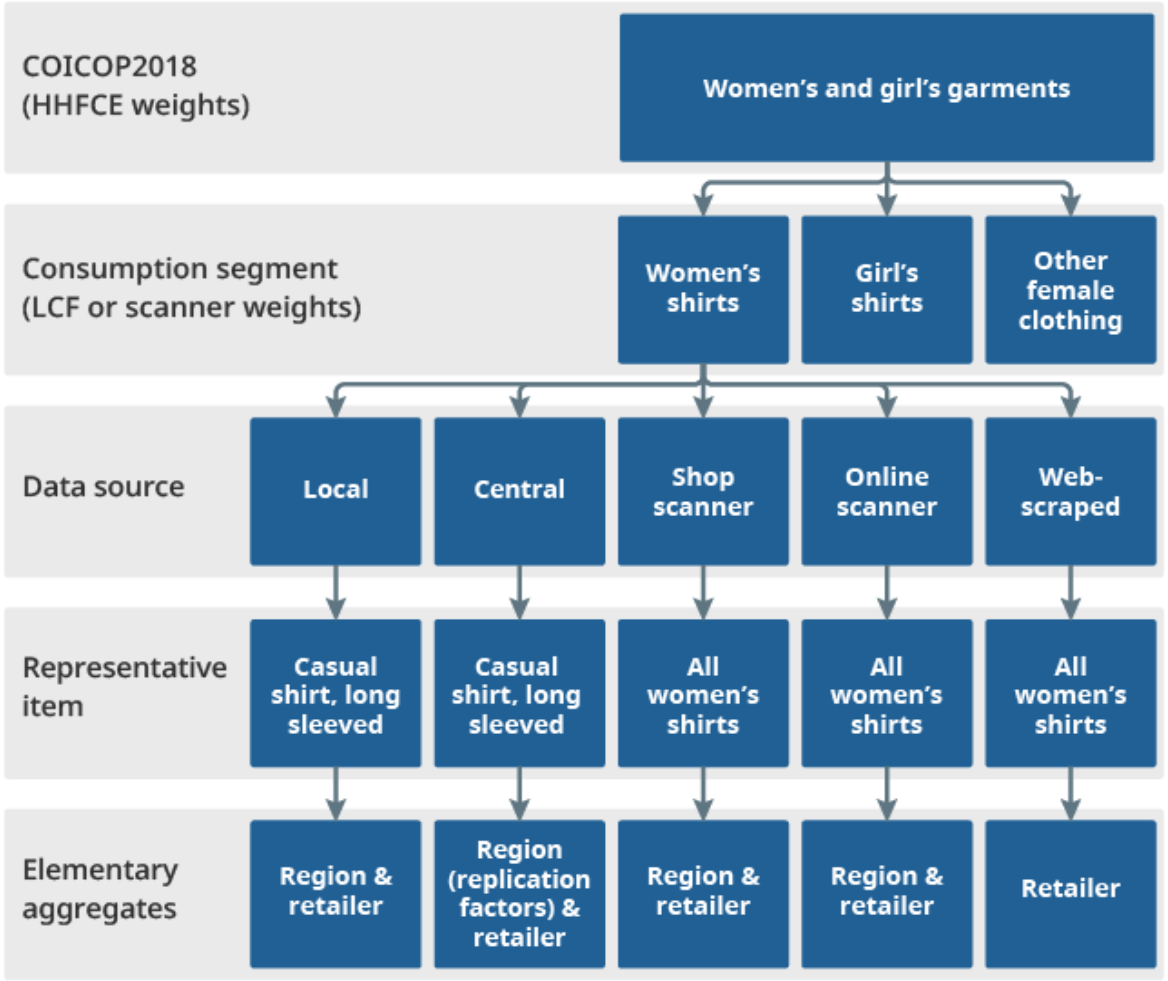
European Commission

Aggregation structures

Examples from the UK, Australia and Belgium

European Commission

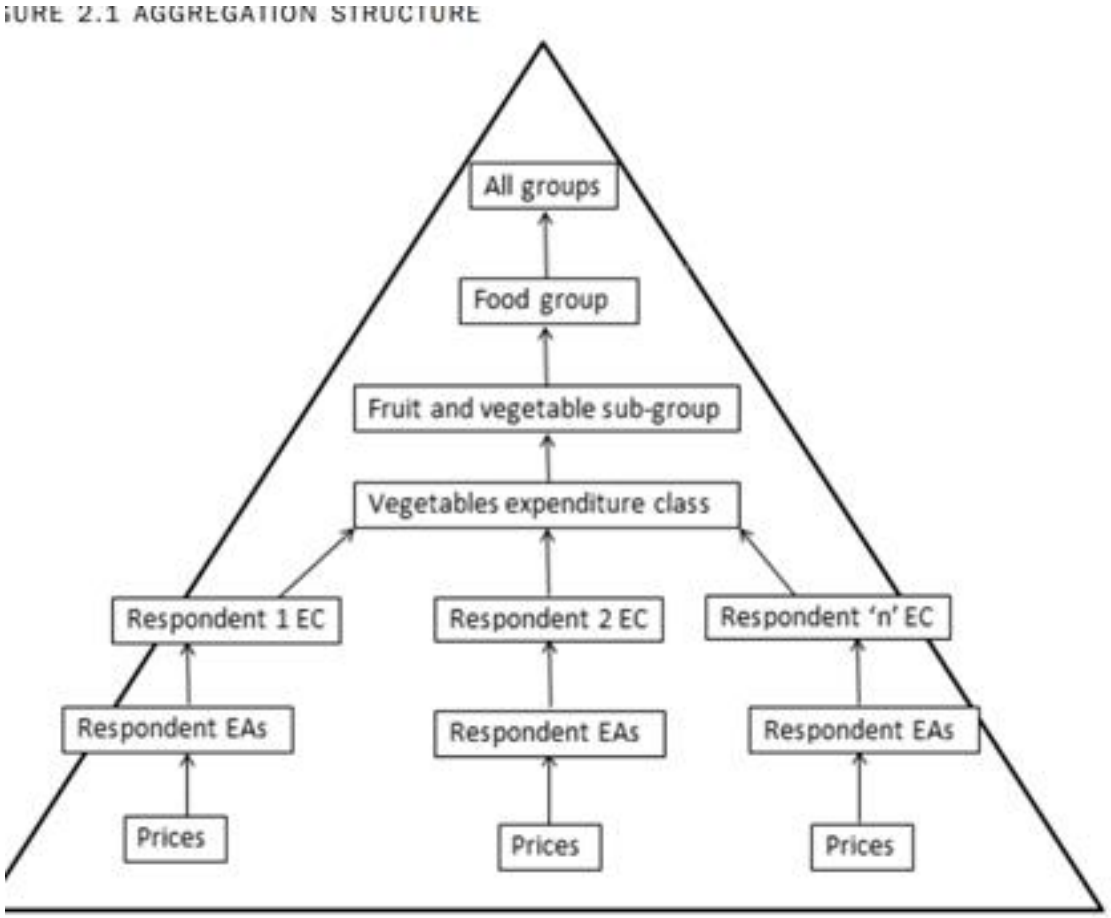
Aggregation structures: UK



- Integration of ADS at the level of a 'consumption segment'
- Local/Central price collection follows the usual approach
- The relatively tight representative items are not used for the ADS
- Stratification of scanner data by 'retailer' and by 'region'; no need for a regional stratification of the web-scraped data

European Commission

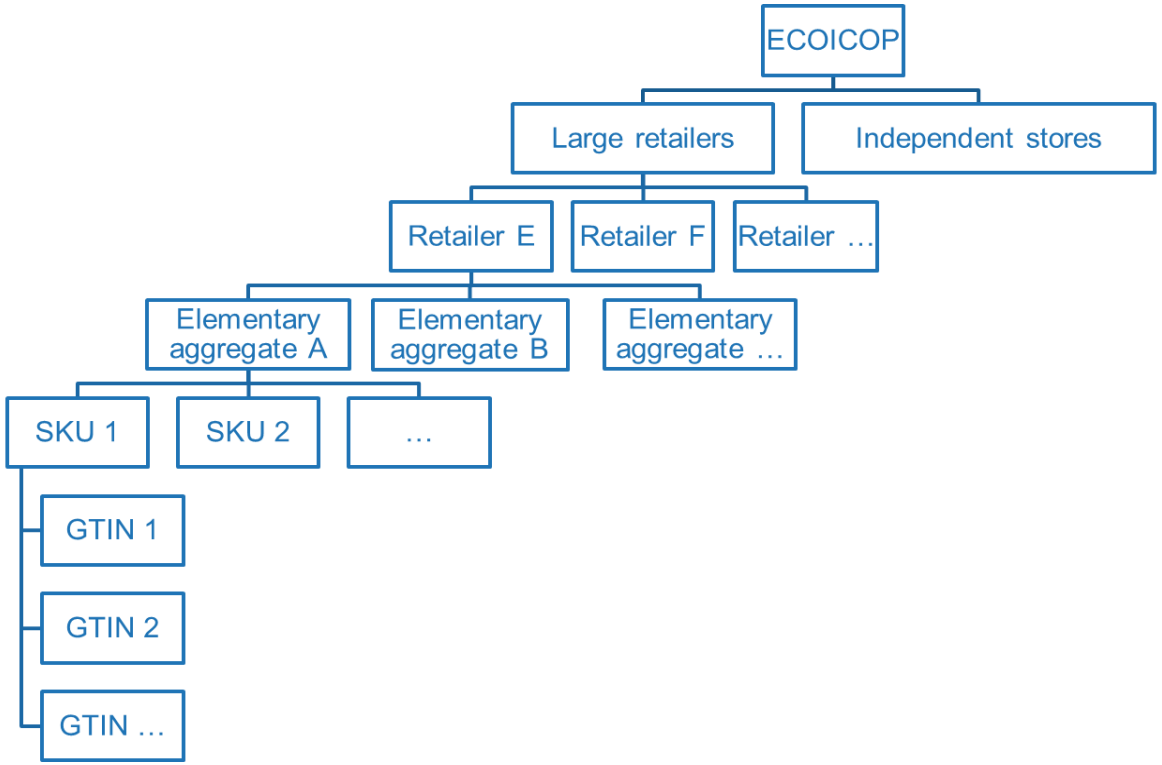
Aggregation structures: Australia



- Integration at the level of the 'expenditure class'
- Stratification by 'respondent', and then by respondent-specific product categories ('Respondent EAs').
- Use of a Törnqvist index to combine the respondent-specific product categories.

European Commission

Aggregation structures: Belgium



- Integration at the level of ECOICOP (5-digit subclass) and type of outlet.
- Stratification of scanner data by ‘retailer’, and then by retailer-specific product categories (‘Elementary aggregates’).

Aggregation structures : Discussion

- The new data sources are separate strata of the target universe for which a weight and a price index must be compiled.

- The level of integration may depend on publication policy, user needs, and the willingness to take into account the increased (product/outlet/regional) coverages of the ADS, the flexible weights of the ADS (for scanner data).

Aggregation structures : Discussion

- What do the **weights of the ‘data source’ or the ‘retailer’** represent? What are the data sources for the weights?
- How should the **price indices within and across data sources be aggregated**?

European Commission

Thank you

© European Union 2020

Unless otherwise noted the reuse of this presentation is authorised under the CC BY 4.0 license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

European Commission