Scanner Data –
some answers, many questions

Berthold Feldmann
Eurostat Unit C5
Price Statistics; Purchasing Power Parities; Housing Statistics

Structure of the talk

1. Scanner data – advantages and disadvantages
2. Main lessons of the last two days
3. Challenges
4. Conclusion

Chapter 1

Scanner data – a new source for various price statistics

Possible data sources

- Traditional survey (price collectors)
  - Declining importance, but will always exist
- Internet prices (manual & scrapping)
  - Very much increasing importance (50% and more)
- Scanner data
  - Increasing, but only slowly
- Administrative data
  - In the larger sense …..

What is scanner data?

Two alternative definitions seem to exist

- SE, NO, NL: quantities and turnover are transmitted by retailers for a large number of EAN
  - ➔ Problem of implicit weighting
- PT: list prices are transmitted by retailers for a large number of EAN
  - ➔ Discounts are missed out?

Advantages

- Optimal coverage (times 20)
- No human error in data
- Productivity gains in data collection
- Information on turnover
- Data can be used for other purposes
  - Purchasing Power Parities
  - Detailed Average Prices
  - Advanced economic analysis
  - **Regional breakdown of CPI/HICP/PPP**

Disadvantages

- Detection of new EAN for the same product is very time consuming
- High investment costs (IT)
- Rules of the legal framework need to be adapted
- Comparability of results to classical survey data?
  - ➔ Unit values can be different from pure prices due to implicit weighting with quantities
- Scanner data only for 20 to 30 percent of all product groups

Chapter 2

Lessons learnt in the Workshop

Some general conclusions

- Only few NSIs use scanner data in daily work, many are however testing the use of it
- Scanner data is a promising source for Multipurpose Consumer Price Statistics (PPP, DAP)
- Scanner data offers high quality information of actual transactions

Practical experience

- The **exchange of views** and experiences between NSIs can save time
  - ✓ Avoid reinventing the wheel
- In scanner data **more volatility** and some large differences to classical survey results
- Major challenge: high **attrition** rate (up to 30%)
- **"Internal" classifications** as links between EAN and COICOP (key) necessary

Relations to retailers

- Look for single point of contact
- Develop written contract
- Remain flexible on format
- Offer tailor made report (evaluation) as reward
- Insist on detailed product characterisation

Chapter 3

Challenges ahead

Weighted or not weighted indices at elementary product level?

- Should quantity information be used at basic calculations?
- Quantities may show huge fluctuations
- In any case, only annual averages should be used as weights
  - Legal obligation!

Further information collected

- Scanner data contains more than prices and quantities
- Product description is important for EAN – COICOP link
- Information on **discounts** is useful for further data analysis for economic purposes

Collect information on what is done in NSIs

- Summing up quantities and turnover over days or even weeks?
- Traditional sample or 80% most sold products?
- Product replacement of individual product offer in consumption segment correct?
- Formula at elementary level?

Chapter 4

Next steps

Enhance transparency

- Update and enrich Eurostat's **overview of current practice** in MS
- Use the **Wiki website** on recommendations for setting up a "road map"
  - Obtaining data
  - Cleaning data
  - Link EAN – COICOP
  - Index formula
- Simultaneously update the Methodological Manual with the recommendations of the road map

Legal obligations

- National law may set different obligations in different countries
- Are stores obliged to supply scanner data?
- Can European legislation help?
- Subsidiarity needs to be respected
- **Next step: carefully formulate appropriate rules in forthcoming Regulation**

Scanner data versus survey data

- Different results from the two alternative sources for consumer price indices
- A thorough analysis is required to explain these differences
- **Next step: Set rules for the permitted compilation of scanner data based HICP**
  - in order to assure comparability

Use a sample or (nearly) all data?

- Current practise differs between countries
- Advantages and disadvantages of the two approaches should be discussed carefully
  - Using a sample is closer to the traditional price observations
  - Using 80% of most sold products might offer new opportunities
- **Next step: A harmonised approach should be achieved**

Link EAN - COICOP

- EAN **differ** from country to country
- A high proportion of products and hence EAN have a **life span** of less than a year
- **Automating** the link of EAN to COICOP level 5 or below is not a trivial task
- In the long run a **harmonisation** of EAN across Europe would be desirable
- **Next step: Create a European repository for mapping EAN to COICOP**

Conclusion

- ✓ We need more information of what is done in MS (transparency)
- ✓ We need rules in order to foster harmonisation (Implementing Regulation)
- ✓ We need recommendations (road map)
- ✓ Dedicated Task Force?

To do list

- NSIs: comments on document by network of experts
- Eurostat: send out new questionnaire on current practise

Thank you for your attention!

Any Questions ?