Scanner data in CPI/HICP

Randi Johannessen, Head of Price Statistics, Statistics Norway
The purpose of the Norwegian Consumer price index (CPI)

- **Multiple**
  - Indicator for cost of living
  - Inflation targeting (CPI adjusted for taxes and excluding energy)
  - Deflator in National Accounts
  - Wage compensation
  - Indexing contracts

- **Harmonised consumer price index (HICP)**
  - For international comparison

- **Core inflation indicators (CPI-AE, ATE, HICP-CT)**
  - Adjusting for taxes and excluding energy
Traditional data collection - questionnaires

• The main data collection method in the Norwegian CPI 20 years ago

• Standardized (paper) questionnaires
  – Pre-defined items from Statistics Norway
  – Specific product description from the shop
  – Price
  – Check boxes for the identification of sales and quality differences and for the availability of the product

• The Norwegian Statistics Act of 16 June 1989
  – Requires businesses to report information needed to provide official statistics
  – May impose compulsory fines
  – High and stable response rate for the Norwegian CPI
Motives for exploring scanner data in the CPI

• Focus on lower response burden
  – High priority in Statistics Norway
  – But also pressure from businesses to deliver data electronically from the head offices
  – About 33 minutes on average per month filling in food questionnaires (1995)
  – The average overall time was about 13 minutes

• Effectiveness
• Improving data quality
• Increasing availability of electronically data
Data collection methods in the Norwegian CPI

- Web questionnaires: 34%
- Scanner data/electronically registered prices: 19%
- Telephone interview: 15%
- Central data collection (Internet): 2%
- Other statistics in Statistics Norway: 30%
Scanner data in the Norwegian CPI

- **First contact with retail chains**
  - 1997
- **Scanner data from state monopoly liquor shops**
  - 2001
- **Scanner data from 1 petrol chain**
  - 2003
- **Full scale scanner data from all retail chains**
  - 2005
- **Scanner data from the 3 largest pharmacy chains**
  - 2010
- **Scanner data from 1 petrol chain**
  - 2012
- **Full scale scanner data from all retail chains**
  - 2013

Based on the CPI expenditure shares

- 0%
- 15%
- 22%
Scanner data

• Item codes are scanned into the cash registers of shops when the items are purchased

• The data is collected from the chains’ headquarters
  – Contains information on price, quantity, specific shop, location, period and description of the item

• The items are identified by;
  – GTIN (Global trade item number), an international retail product code
  – PLU codes; internal chain specific codes

• Average transaction price (sales/quantity) paid per item

• Cover 3 weeks of the month (food and non-alcoholic beverage)
Scanner data - coverage

• Scanner data from all 3 retail chains → 96 % coverage of market
  – Sample: around 700 000 records each week (200 retail shops)
  – Food, non-alcoholic beverage, beer, tobacco, household cleaning products, food for domestic animals, soil and nutrient for plants, magazines, personal hygiene products

• Scanner data from the state monopoly of liquor store → 100 % coverage
  – Around 20 000 records three times a year

• Scanner data from 3 out of 4 pharmacy chains → 85 % coverage of market
  – Sample: around 450 000 records each month (120 pharmacies)
  – Alcoholic beverage

• Scanner data from 4 of 6 petrol chains → 85 % coverage of market
  – Sample: around 40 000 records each month (70 petrol stations)
  – Engine fuel and some foodstuff like hotdogs, pizza slices as well as tobacco and daily necessities
Methodology

• Initially we replaced prices from questionnaires with scanner data prices on predefined items
  – Requires manually work by lining the GTIN code to our predefined items
  – Still do that for non-food items from retail chains
  – Challenge: high attrition rate of items (GTIN) due to end lifecycle items

• More automatically production system is implemented for food, non-alcoholic beverage and medical products
  – Utilize approximately all the records
  – Matching each GITIN every month
  – Extreme price changes are excluded
Data delivery

• Scanner data sent as attachment in e-mails
  – Some are zipped (password protected) and others not
  – Some send an automatically generated emails with the data

• Size of data files differ among the providers
  – From 10 to 42 000 kilobytes (KB)

• Other (preferred) ways of reporting scanner data:
  – MOVE-it - a secure and efficient channel - a secure web site where providers log into, upload and send their files
  – SFTP (secure file transfer protocol) - a secure way to transfer files between local and remote servers where providers can drop the data automatically
Scanner data - quality improvements

• Actual transaction price for each item (sales/quantity)
  – The risk for manually reported errors connected to questionnaires is eliminated
  – Less dependent on the shop’s motivation for reporting prices
  – If using price collectors – avoid shelf prices that may not always be correct

• Larger data coverage
  – Approximately full scale item coverage

• Increased response rate

• Opens up for possibilities for more advanced quality adjustment and calculation methods
  – Utilize both price and quantity information
Scanner data – quality improvements

• However;
  – Few data suppliers makes the statistical agency more vulnerable if they fail to deliver the data
  – Emergency plans - what to do if data is missing
  – Without the right IT-system, the amount of data may appear unmanageable
Efficiency gains?

• Large establishing costs for recipient of the data
  – Time consuming process – negotiation with the chains
  – Testing and analysing the data is necessary
  – IT systems and technological solutions must be established
  – Written contracts and strict time schedules
  – No changes in formats when implemented

• Minor cost for chains once scanner data is establish

• Lower response burden on businesses
  – No questionnaires or monthly visit of price collectors

• Less use of staff in the data collection unit when producing the monthly questionnaires
Efficiency gains in the Price statistics unit?

• Increased the time spent on producing the CPI/HICP the last 15 years – but also improved quality of CPI/HICP

• The increase is a result of several factors:
  – several new indicators for underlying inflation, such as the CPI-AE and CPI-ATE
  – the Rental Market Survey
  – expanded CPI/HICP – more council regulations
  – the CPI/HICP covers more consumption groups
  – more complex production process due to various data sources
Concluding remarks

- Reduced response burden on businesses
- Improved CPI in terms of data accuracy and coverage
- Less uniform and more complex CPI production system
- Important to make the chains deliver scanner data on standardized form
- Important to establish good routines and good relationship with the chains
- Acceptance of scanner data as an important data source
  - Not a source that should be limited to CPI/HICP
  - Data can be used for different statistical purposes (PPP, Index of retail sales), as well as in research
- In some countries access to scanner data is more complex
  - Pay for the data?
  - Market research companies and not directly from the chains?
Further plans

• Expanded use of the non-food scanner data from retail chains
• Introduce scanner data in other consumer areas
  – Electronic appliance
  – Clothing and footwear
  – Sports equipment
• More secure delivery of scanner data
• More streamlined production system
  – Data warehouse?
• Provide information to link GTIN automatically to consumption groups in CPI
  – Utilize the chains own classification codes
• More use of data from the internet (web-scraping)