

# The use of online prices in the Norwegian CPI

**Ottawa Group Meeting  
Tokyo, 20-22 May 2015**

Ragnhild Nygaard  
Statistics Norway

# Outline

- Online data project
- E-commerce in Norway
- Online prices in the CPI/HICP
- Web scraping
  - Import.io
- Test calculations for personal care products
- Data collection of airline fares

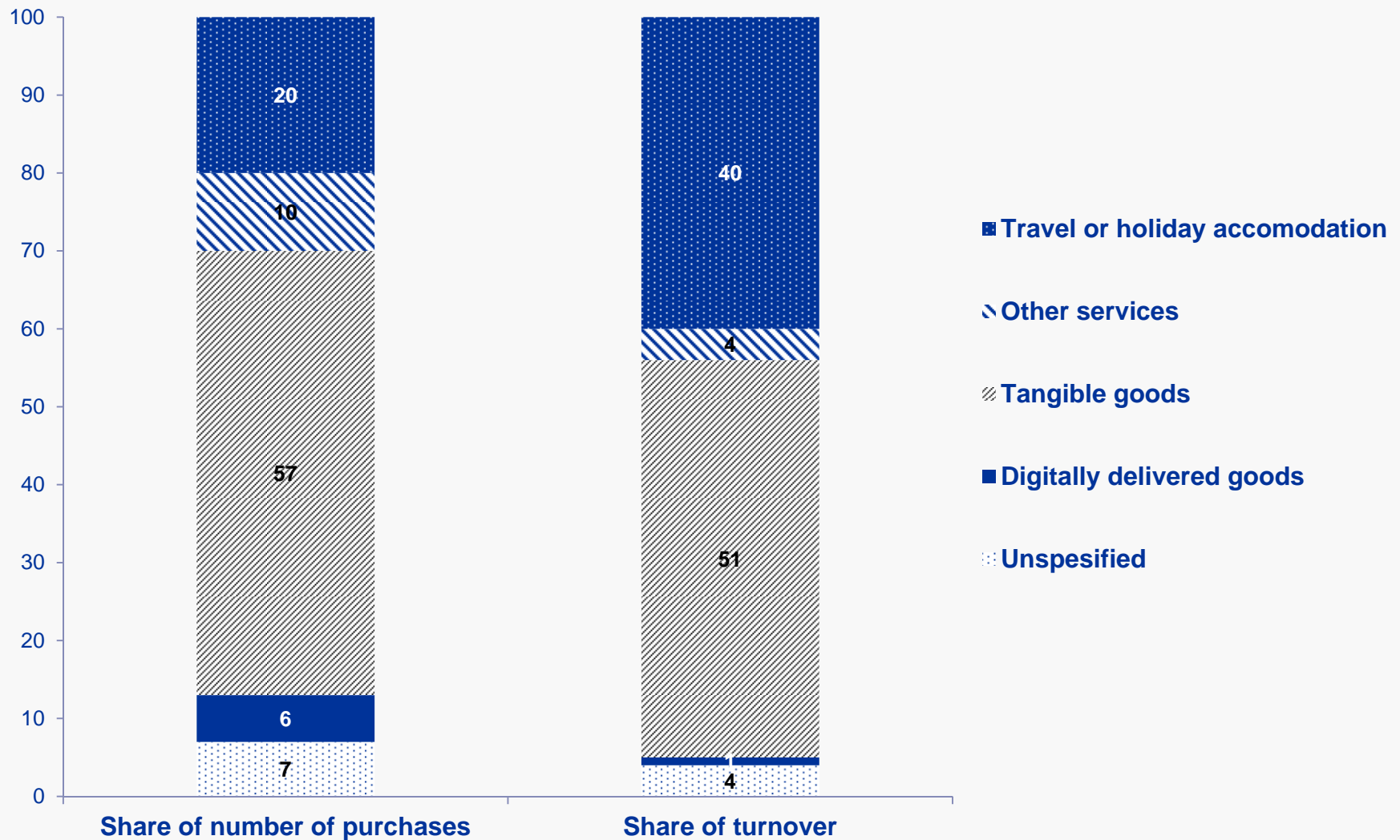
# Online data project

- Project partly financed by Eurostat
  - “Multipurpose Price Statistics”
- Several other NSIs working on online data and web scraping
- Eurostat guidelines
  - Online purchases shall be included in the HICP according to their significance
  - Shall cover online purchases of goods available for purchase on the economic territory of the Member State, regardless of the residential status of the seller
- Two-sided aim:
  - Increase the amount of online prices in the Norwegian HICP/CPI and improve the mix of online prices and traditional store prices
  - Create an efficient online data collection

# E-commerce in Norway

- The Norwegians among those who shop the most
  - High living standards
  - High internet availability (approx. 96 per cent)
  - Outstretched country, stores with limited product range
- The traditional retail stores still maintain a very strong position
  - Online purchases estimated to less than 10 per cent of the total turnover
- Goods that Norwegians buy online are e.g. home electronics, clothing, books and personal care products
- Travel/holiday accommodation makes out approx. 90 per cent of all purchases of services
- Foreign shopping
  - Estimated to be around 23 per cent of the total online purchases of tangible goods

# Goods and services purchased online



# Online data collection in the Norwegian CPI/HICP

- Centralized data collection
  - Mostly services
  - About 18 per cent of total CPI basket measured by CPI weight shares
  - Internet the leading purchaser channel
    - ♦ Airline fares
  - Administratively determined prices/National prices
    - ♦ Health services etc.
    - ♦ Streaming services
- Manual data collection
- Store sample based data collection
  - Mostly tangible goods

# Web scraping



- Tested different software
  - Import.io, iMacros
  - Free software which can easily be downloaded from the internet
- Advantages;
  - No programming skills/no previous knowledge required
  - Easy to make adjustments
  - Useful illustrations and examples online
  - “Cloud” storage, safe storage of data
- Disadvantages;
  - External software outside the control of the NSI
  - Other technical solutions may be establish in the long run?

## Web scraping in practice

- Daily download of prices and price related information from 4 online retailers registered in Norway with highest expenditure figures
  - Personal care products and home electronics
- Data downloaded and “cleaned” into csv-files and then further loaded into SAS for analysis and calculation
- Some obvious efficiency gains compared to sending web questionnaires to online retailers, but are there potential quality improvements?



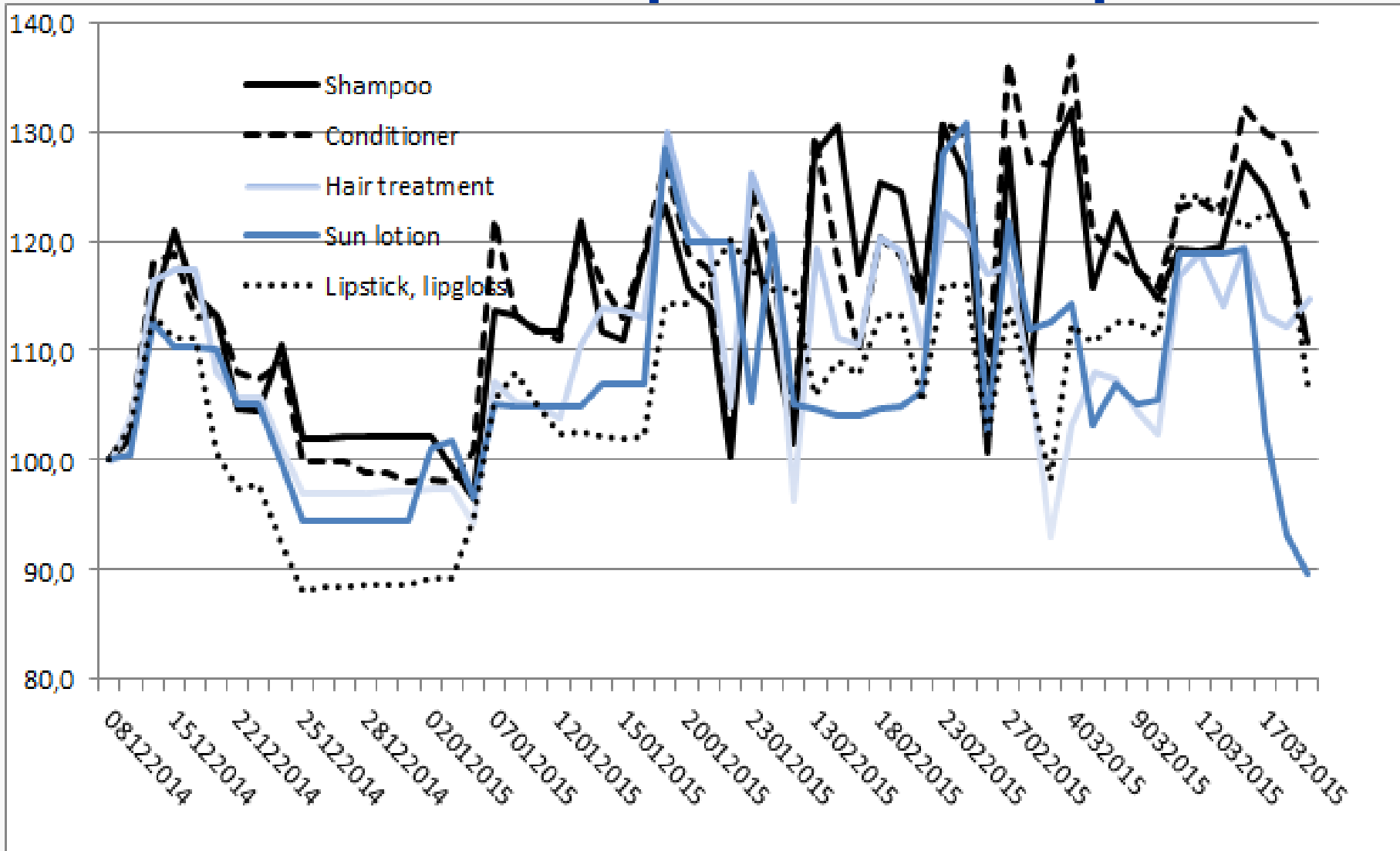
# Test calculations - personal care products

- The way we calculate the indices and the assumptions we make may strongly affect the outcome
- Volatile price movements
  - Daily changes in prices
  - Most likely more volatile compared to physical store prices
  - Dynamic pricing
    - ♦ Prices may change according to product's popularity online
- How to use the large amount of data available without any expenditure shares?
  - Only extract data from stores with the highest expenditure shares
  - Only the most sold products
  - Equal weighting

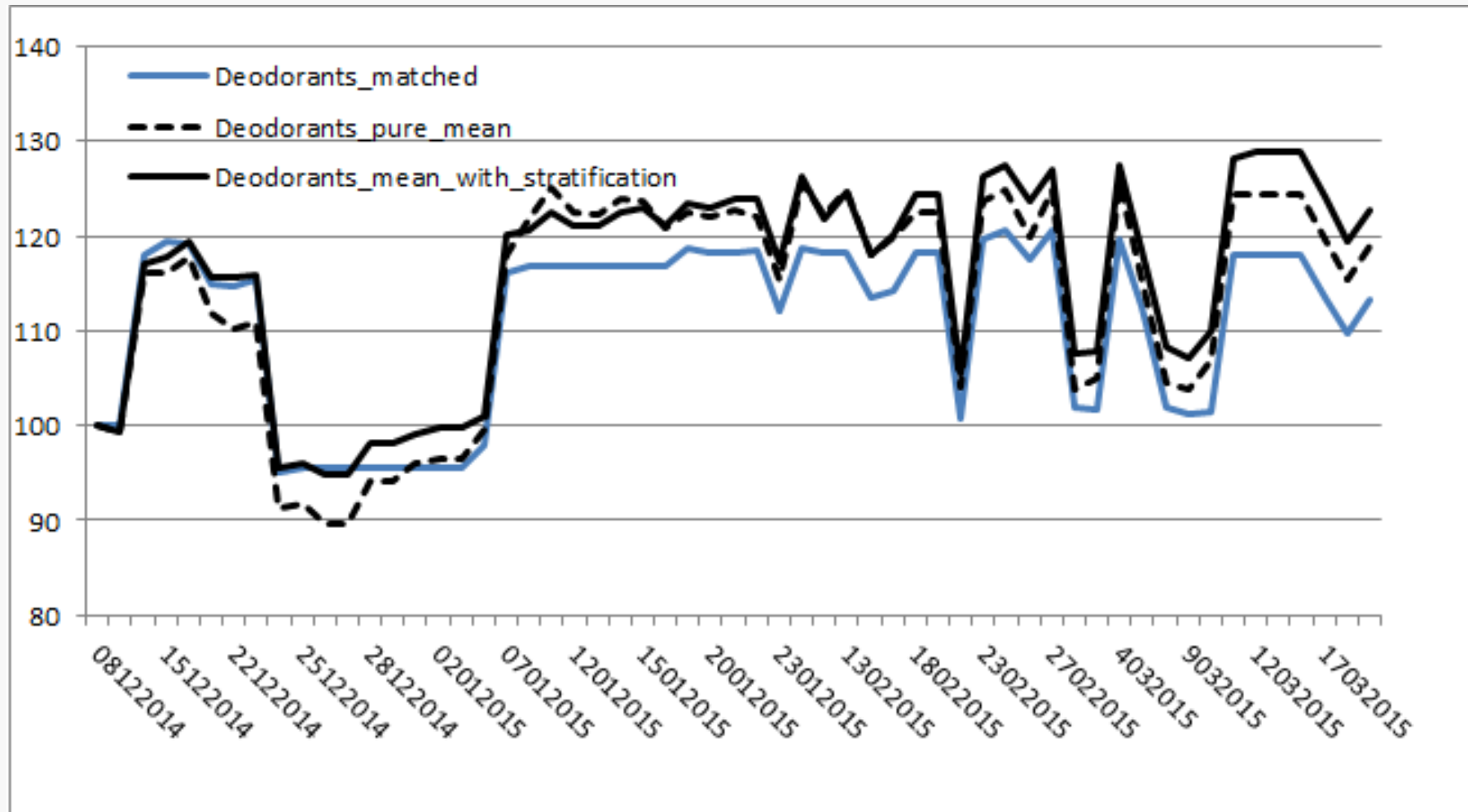
## Test calculations, cont.

- Use the URLs to classify the products
  - Create detailed product groups
- Use the product description to extract the unit specification
  - E.g. 50 ML, 1000 ML etc.
- Calculations
  - Daily chained indices based on matched samples
  - Daily chained indices based on pure average prices
  - Daily chained indices based on average prices with stratification

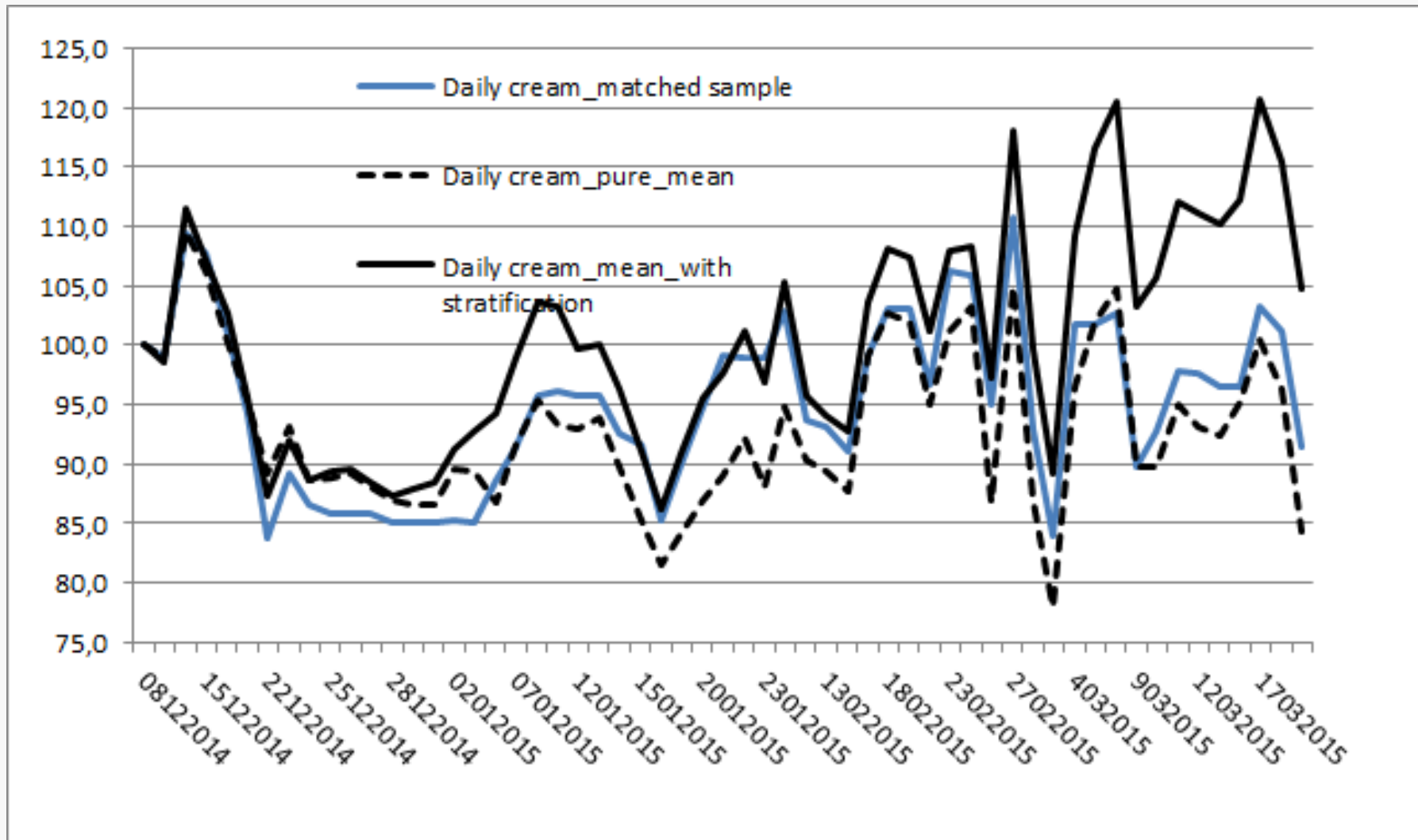
# Test calculations – personal care products



# Test calculations, cont.



# Test calculations, cont.



# Data collection of airline fares

- Manual, time consuming data collection
  - “Copy – paste” solution
- Other technical solution
  - Go directly to travel search engines and not through web sites
  - Even small adjustment in the visuals of a website, might cause empty extraction
- Google QPX Express
  - Allows 50 requests per day free of charge
    - ♦ No room for increasing the coverage
  - Violate the terms of conditions?
  - The airline companies decide which prices to be displayed
    - ♦ Differences in price
  - Need programming skills

## Data collection of airline fares, cont.

- Airline companies' own search engines?
  - Extract data directly from their own search engines
  - Might use the Norwegian Statistics Act
  - No response burden for the data provider