Alternative data sources for price statistics: Statistics Norway

KJERSTI NYBORG HOV

WEBINAR 1 JULY 2025

UN REGIONAL HUB FOR AFRICA

Statistisk sentralbyrå
Statistics Norway

Kjersti Nyborg Hov

- MSc in Economics and Business Administration

- Senior advisor at Statistics Norway, Division for Price Statistics

  - CPI development projects, alternative data sources and other improvements for the CPI

  - International CPI development projects and technical assistance, mainly in Africa and Asia
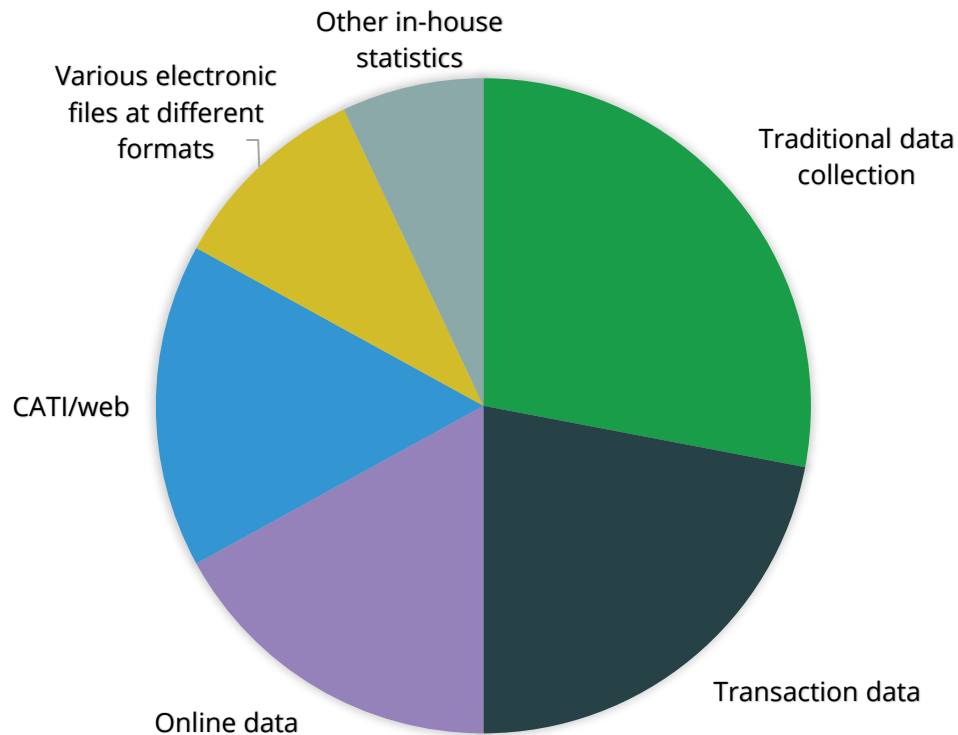
✉ Kjersti.Nyborg.Hov@ssb.no

▢ (+47) 976 05 808

Statistisk sentralbyrå
Statistics Norway

Statistics Norway

Alternative data sources are increasingly available and utilized for price statistics

- ○ Globalization

- ○ Growing e-commerce

- ○ Increased product range

- ○ Reducing data collection costs

- ○ Increasing quality of CPI measurement

Statistisk sentralbyrå
Statistics Norway

Alternative data sources

Web scraping

- Price information only

- Large amount of metadata often available

- Potentially unstable
  - Web site might change
  - Web scraper blocked from the web page

API / pre-defined datasets

- Price, but rarely turnover information

- More stable than web scraping
  - Structured data
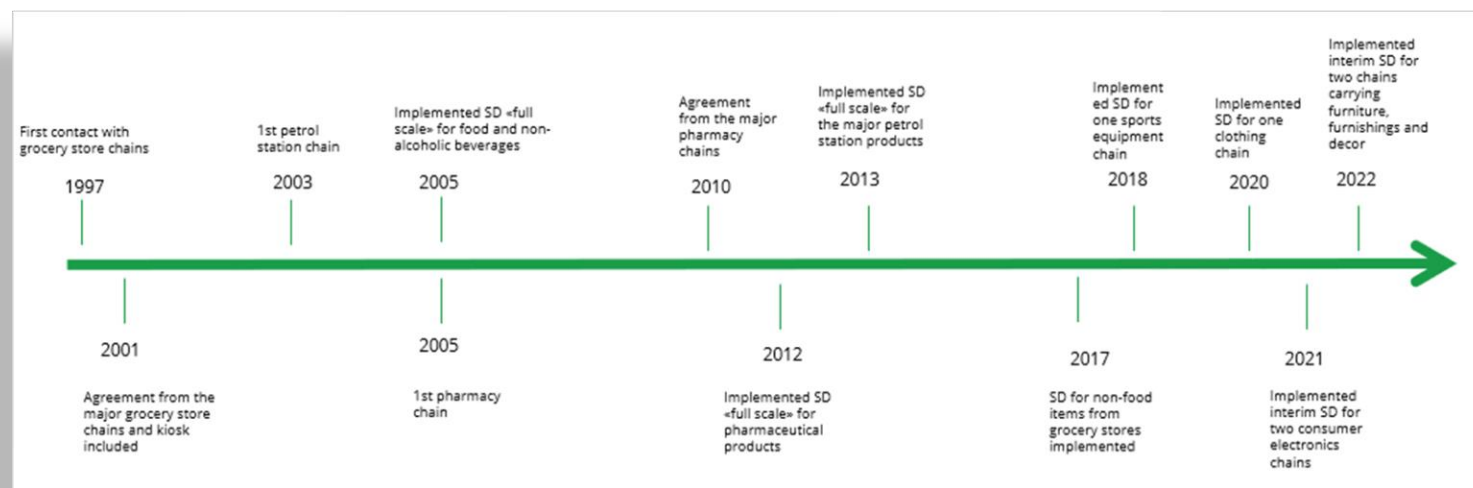 - Less changes to the structure

Transaction data

- Price and turnover information

- Structured and stable data source

The preferred data source

Statistisk sentralbyrå
Statistics Norway

Statistics Norway

- A step-by-step approach: from a *static* to a *dynamic* approach

  ◦ I.e. *picking price observations* for specific item varieties → *flexible item basket* with increased data coverage

- Transaction data in use "full scale" for COICOP 01 since 2005, <u>but</u> methodology behind has changed over time

  - However, always by matched items (outlet*GTIN) over time, for food..

    ..for other product groups: other solutions

First contact with grocery store chains — 1997

1st petrol station chain — 2003

Implemented SD «full scale» for food and non-alcoholic beverages — 2005

Agreement from the major pharmacy chains — 2010

Implemented SD «full scale» for the major petrol station products — 2013

Implemented SD for one sports equipment chain — 2018

Implemented SD for one clothing chain — 2020

Implemented interim SD for two chains carrying furniture, furnishings and decor — 2022

Agreement from the major grocery store chains and kiosk included — 2001

1st pharmacy chain — 2005

Implemented SD «full scale» for pharmaceutical products — 2012

SD for non-food items from grocery stores implemented — 2017

Implemented interim SD for two consumer electronics chains — 2021

Statistisk sentralbyrå
Statistics Norway

How to approach retailers

- Web scraping netiquette..

  ◦ Even if it is legal to web scrape, you should always identify yourself as the NSO

  ◦ For stability and security reasons an agreement with the retailer is advisable

  … can be used to get access to more stable data? Even transaction data?

- Retailers that provide data to other NSOs?

- 3rd party providers?

- Statistical act?

  ◦ … a mutual understanding and agreement is advisable

Statistisk sentralbyrå
Statistics Norway

Which retailers to approach

- CPI relevance

  - Weight in the CPI-basket?

  - Product group homogeneity – do we need larger item variety coverage?

- Analyze market shares of the retailers for a product group

  - Many small retailers or a few large retailers dominating the market?

- How representative are the retailer(s) of current consumption pattern?

  - Supermarkets vs open markets

  - Combine data sources?

Statistisk sentralbyrå
Statistics Norway

Analyze the type of product

Item churn? Life cycle effects? Large quality differences? Rapid technical improvements?

Food, beverages, pharmaceutical products, fuel for cars..

- (relatively) stable item codes, no evident life cycle effects

Linking item over time at **item code level** (GTIN or similar unique item code)

Clothing and sport equipment

- Large degree of item churn

- Life-cycle pricing

Linking **groupings** of individual items of similar type and quality over time (unit value)

Consumer electronics

- Large degree of item churn

- Life-cycle pricing

- Rapid technological improvements

..will also need **quality adjustment methods**

Statistisk sentralbyrå
Statistics Norway

Thank you for your attention

Takk!

Statistisk sentralbyrå
Statistics Norway