What micro price data teach us about the inflation process:
Webscraping in PRISMA

Chiara Osbat* (ECB, Directorate General Economics)
Based on contributions to the Price Setting Microdata Analysis (PRISMA) network.

The views expressed in this presentation are mine and do not necessarily reflect those of the ECB or the Eurosystem

EUROPEAN CENTRAL BANK
EUROSYSTEM

The Return of Inflation
OeNB, 24/05/2022

PRISMA: the **PR**ice-**S**etting **M**icrodata **A**nalysis **N**etwork

Research Network by the European System of Central Banks to deepen the understanding of price-setting behaviour and inflation dynamics in the EU, gaining new insights into a key element of monetary policy transmission. Collects and studies various kinds of microdata:

1. Official data underlying official price indices (CPI and PPI)

2. Scanner data

3. Online prices

The advantages and disadvantages of online price data

✓
- **High Frequency:** daily, and in principle can be scraped intra-daily to study dynamic pricing
- **Timely**: yesterday's data are available today
- **Precise:** the price of a given product, including metadata (e.g. is it discounted)
- **Large volume** ☺

✗
- **Not always representative:** can refer to a specific zip code, or specific to customer
- **Non-homogeneous**: Each website contains different information
- **"Bulk":** Must be classified and mapped to statistical categories
- **Large volume** ☹

Just a few clicks away? The challenges of building web-scraping data pipelines

Scraping is the easy part!

- **Conceptual Definitions**:
  - Creating a harmonized data model
  - Developing meaningful aggregation rules
- **Collecting, validating and storing the data and metadata**
- **Monitoring the daily data flow**
- **Classifying the data** to a common classification system (ECOICOP)
  - Analytical aspect: developing AI classifiers – multilingual problem!
  - MLOPS aspects of continuously developing, monitoring, maintaining, and deploying the classifiers
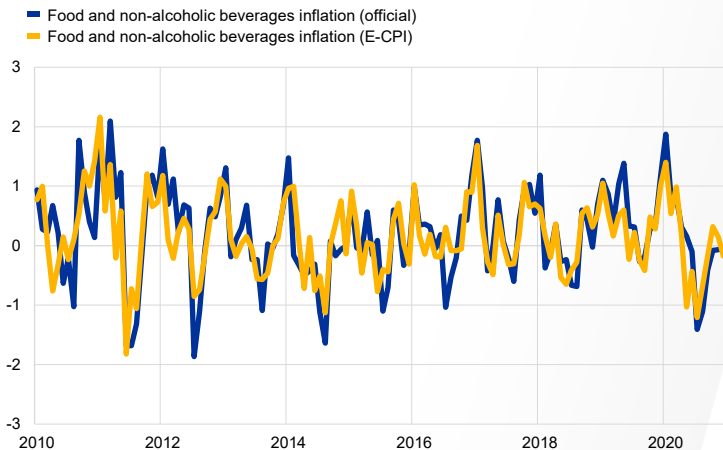
The pipeline must be automated to be sustainable!

But how can we use web-scraped data?

Big Data Pipeline on AWS
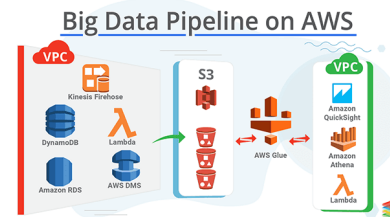
Nowcasting CPI: methodology matters!

Comparison of official CPI and online prices
(Based on official product selection and aggregation, m-o-m %)

- Food and non-alcoholic beverages inflation (official)
- Food and non-alcoholic beverages inflation (E-CPI)

Source: Macias, Stelmasiak and Szafranek.

Macias, Stelmasiak and Szafranek (IJF 2022)
"*Nowcasting food inflation with a massive amount of online prices*":

- Substantial and statistically significant reduction in the nowcasting errors with respect to popular benchmarks.
- The volume helps…
- But a lot of work must go into ECOICOP classification, choice of product and precise application of the official CPI methodology
- During 2020 the accuracy of the baseline model increased with respect to the benchmark

Inflation measurement: online data are a lab for experimenting with methodology

Goldhammer, Henkel, Osiewicz (2021) "Bias related to the Bridged-overlap- and Link-to-Show-No-Price-Change Method"

Test calculations for different quality adjustment methods
(Index, 27/08/2018 = 100)

Legend:
- Dutot - BO
- Dutot - LNP
- Dutot - DPC
- Jevons - BO
- Jevons - LNP
- Jevons - DPC
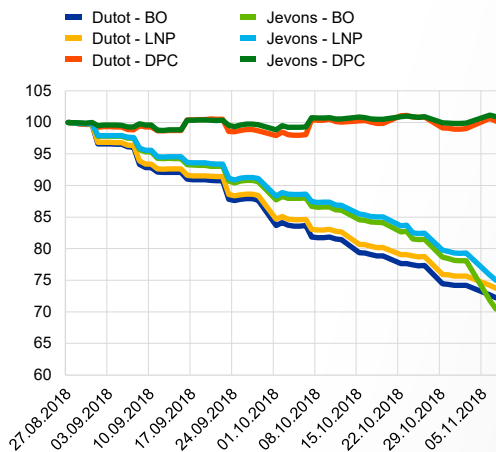
Source: ECB calculations. Notes: DPC: direct price comparison; LNP: link-to-show-no price change; BO: bridged overlap.

Three implicit methods to account for quality changes:

- Direct price comparison (no quality difference)
- Link to show no price change (the price difference is equal to the quality difference)
- Bridged overlap (the price difference equals the average percentage price change of all observed products in the same or similar product category),

Disregarding price changes at replacement can lead to a downward drift in a price index.

Very important: check the underlying assumptions!

# Inflation monitoring: online data are the only source for following special events in real time

VAT change in Germany: month-on month average price change

Source: ECB staff calculations.
Notes: Web scraped data from German online supermarkets, containing information mainly on food, beverages and personal care items. Data are collected daily. The chart shows the daily unweighted average of 4-week price changes. 4-week price changes are calculated as the percentage change of the price of a product on a given day compared to the price of the same product on the same weekday four weeks before. Latest observation: 26 July 2021.

VAT change in Germany: frequency of price increases and decreases
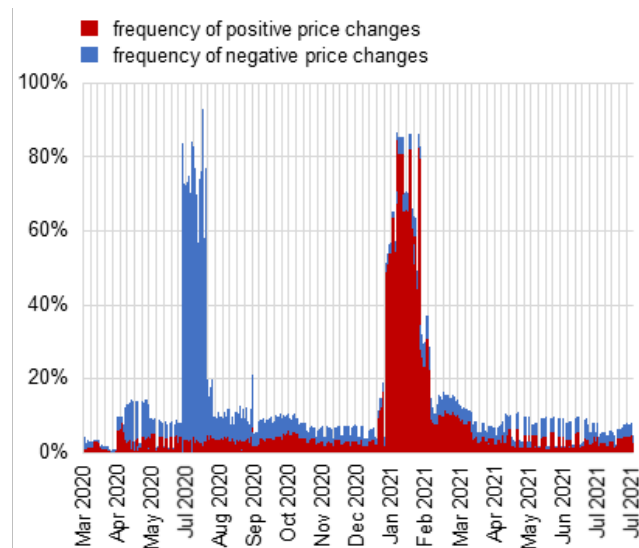
Source: ECB staff calculations.
Notes: Web scraped data from German online supermarkets, containing information mainly on food, beverages and personal care items. Data are collected daily. The chart shows the daily share of products that experienced a price change compared to four weeks before. Latest observation: 26 July 2021.

Inflation monitoring: daily data give insights on what is happening with the "return of inflation"

Germany: frequency of price increases and decreases during VAT change and in 2022
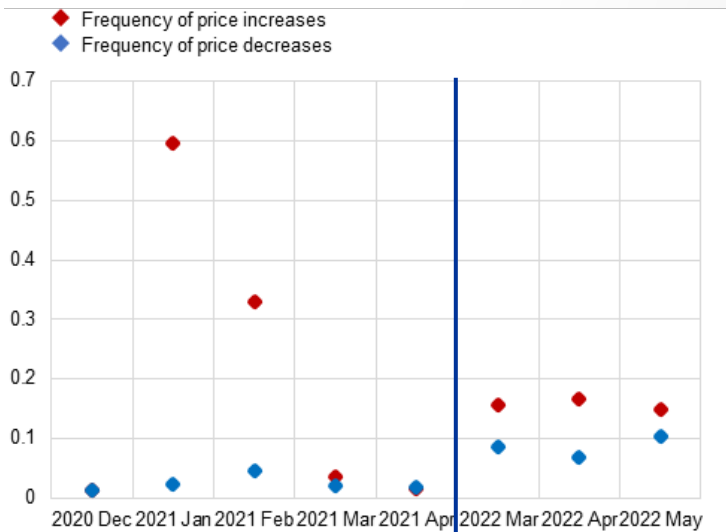
Source: ECB staff calculations based on daily online price data.
Notes: Web scraped data from a German online supermarket. Products on sale are excluded. Latest observation: 16 May 2022.

Germany: size of price increases and decreases during VAT change and in 2022
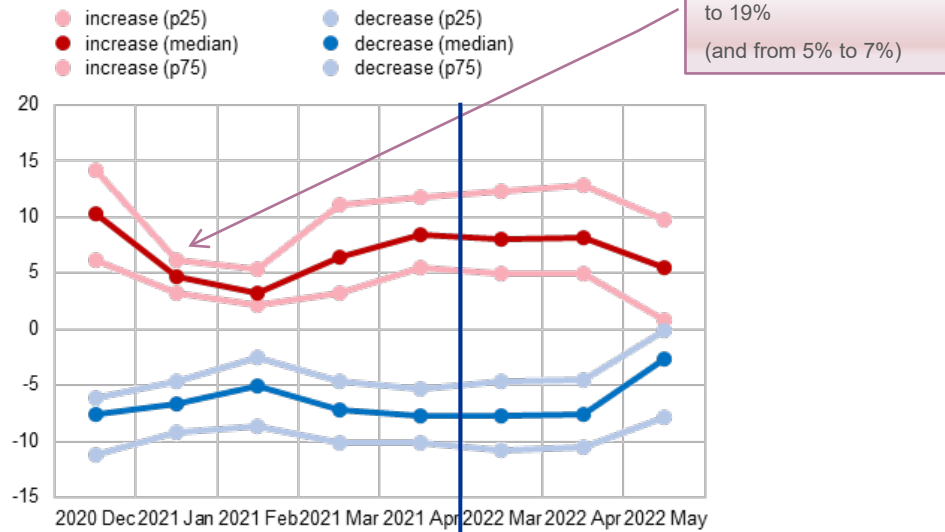
VAT returns from 16% to 19%

(and from 5% to 7%)

Source: ECB staff calculations based on daily online price data.
Notes: Web scraped data from a German online supermarket. Products on sale are excluded. Latest observation: 16 May 2022.

Online data are especially useful if complemented with scanner data!

Looking at change of behaviour from 2021 to 2022, we found heterogeneity across retailers.

What explains this heterogeneity?

Does it depend on market power? → This is a central question in understanding inflation dynamics.

This is where the complementarity between online and scanner data shines:

- Online data are very timely and can point to fresh facts

- Scanner data are "late" but much richer and can help to deepen the study of heterogeneity in pricing behaviour:

    - Estimate price elasticities, consider impact of market shares.

    - Household panels help to understand heterogeneity of experienced inflation across different demographics: Very relevant for policy when inflation returns strongly on energy and food!