TALL
SOM FORTELLER

New ways of measuring price development on consumer electronics

Division for Price Statistics
Statistics Norway

17th Ottawa Group Meeting
Rome, 7-10 June 2022
Session 3

Study done by Kjersti Nyborg Hov,
Gunnar Larsson and Ragnhild Nygaard

Statistisk sentralbyrå
Statistics Norway

Background

- Statistics Norway uses scanner data for main areas like groceries, clothing, pharmaceutical products and fuels

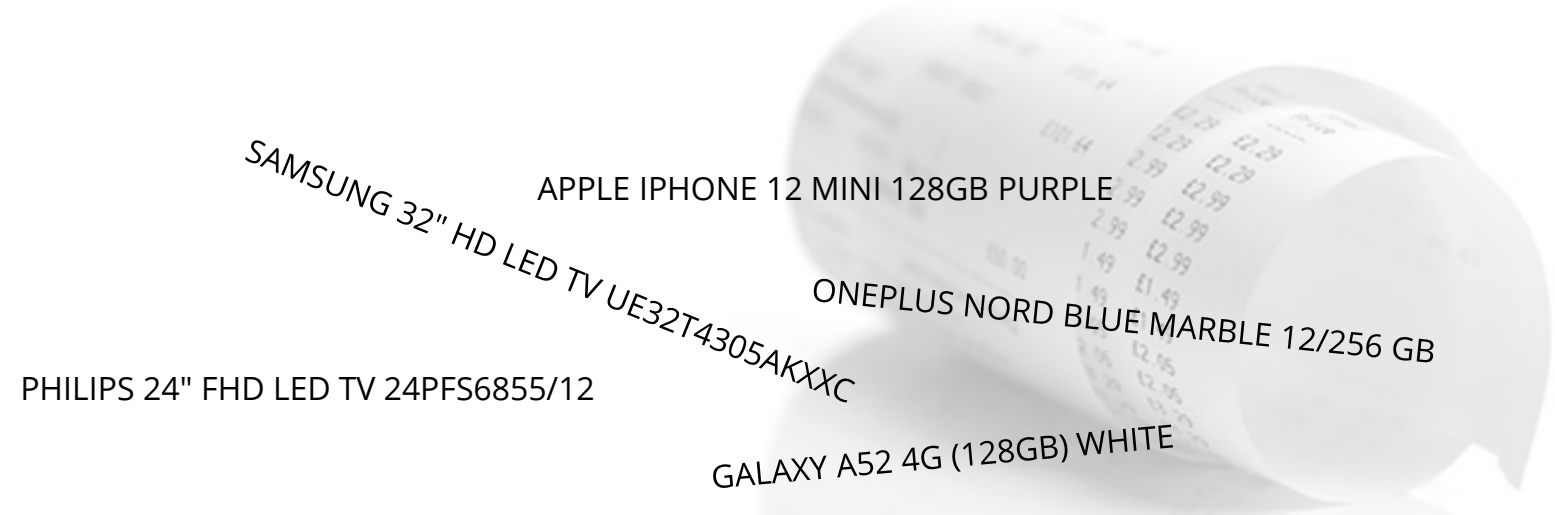
  → Wants to expand the use of scanner data to consumer electronics

- Consumer electronics poses additional challenges
  - High item churn, technological development, rapid quality changes
- Therefore: Need of quality assessment and adjustment

- Eurostat grant project
  - Find practical cost-efficient solutions for using scanner data for consumer electronics in “large scale”

Scanner data on consumer electronics

- Regular data transmission from the two leading retailers on the Norwegian market
- Weekly data delivery, covering both physical and online stores
  - Data aggregated to chain level and type of store
- Covers the whole range of products directed towards consumers
- Short item description provided
  - Note that the most important item characteristics often are embedded in the text

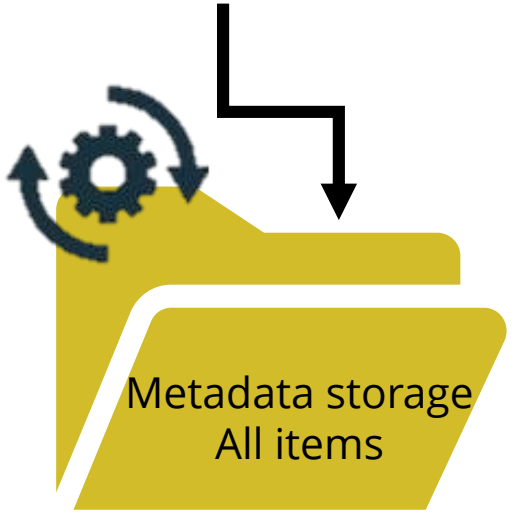


Web scraping metadata - weekly

New items

productType	itemName	week	year	lenke	Produsent	erandøren	Materiale	rimærfarg	tall kamer	Megapiksle	otofo	ygge	oppl	destseåpne	lef	Paneltype	rmoppløs	touchskjer	rikseltetthe	jermstørre	RAM
Mobiltelefon	GALAXY A13 64 GB S	16	2022	https://wv	Samsung	SM-A135F	Plast	Svart	4 Stykk	8 MP					Ja	TFT	720x1600	Ja		6.6 Tomm	4 GB
Mobiltelefon	SAMSUNG GALAXY A	16	2022	https://wv	Samsung	SM-A536B	Glass	Svart	4 Stykk	32 MP	Ja	Ja	Ja	f/2.2		Super AMOLED		Ja		6.5 Tommer (inch)	
Mobiltelefon	GALAXY A13 64 GB B	16	2022	https://wv	Samsung	SM-A135F	Plast	Lyse blå	4 Stykk	8 MP					Ja	TFT	720x1600	Ja		6.6 Tomm	4 GB

- Assumption: In general the same items sold in the two competing retailers, hence scraping only one retailer
  …Only partially true
- The technical solution set up and maintained by CPI staff



Statistisk sentralbyrå
Statistics Norway

Metadata online

- Overwhelming amount of information
  - Information varies across different product groups
    - 25 – 100 variables per item
  - Several variables explain similar attributes
    - *Battery capacity on the phone* vs. *Battery capacity on standby*
  - Metadata suffers from inconsistency and incompleteness

- No standardization of the technical specifications
  - Resource-intensive, especially aiming at monthly production

Vekt og dimensjoner

Nettomål uten emballasje (D x B x H)	309 x 205 x 17 mm
Bruttomål med emballasje (D x B x H)	523 x 72 x 306 mm
Nettovekt	1.3 kg
Bruttovekt	2.162 kg

Design og utforming

Primærfarge	Sølv

Operativsystem

Operativsystem (familie)	Windows
Operativsystem (versjon)	Windows 10
Operativsystem (bit)	64-bit

Prosessor

Prosessormerke	Intel
Prosessor (familie)	Intel® Core™ i
Prosessor (generasjon)	11th Gen Intel
Prosessor (modell)	1135G7
Antall kjerner	4 Stykk
Prosessorhastighet	2.4 Ghz
Prosessorhastighet med boost	4.2 Ghz

Grafikk

Dedikert grafikkprosessor	Nei
Grafikkort Merke	AMD
Grafikkprosessor (familie)	Radeon™ Vega
Grafikkprosessor (modell)	Radeon™ Vega 8 Graphics

Lagring

Type lagring	SSD
Total lagringskapasitet	512 GB
SSD kapasitet	512 GB

Arbeidsminne (RAM)

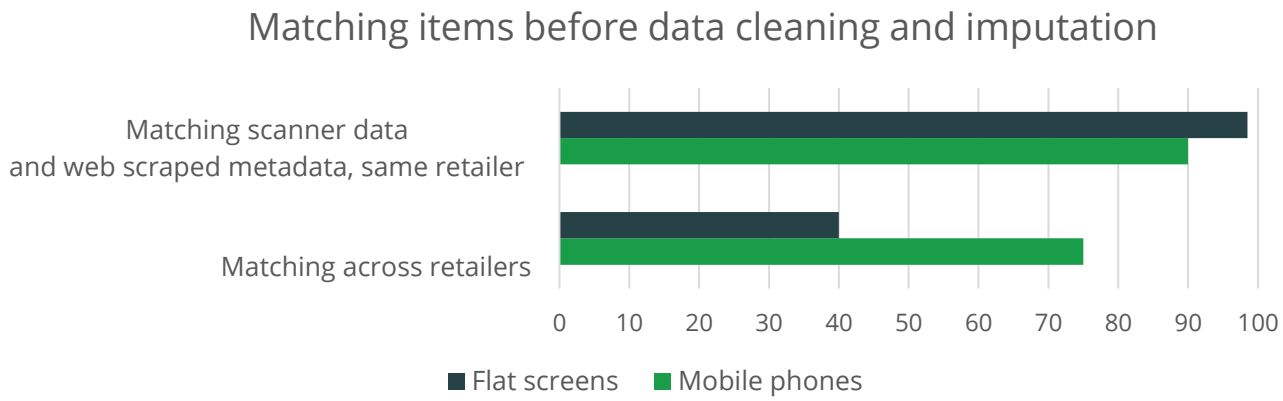
Minnekapasitet	8 GB
Type RAM	DDR4
RAM	8 GB (2x4 GB)
Antall RAM-spor	2

Skjerm

Skjermstørrelse	15.6 Tommer (inch)
Skjermoppløsning	1920x1080 (Full HD)
Paneltype	LED
Skjermteknologi	LCD
Touchskjerm	Nei

Combining data sources

- Large differences in number of matches of items transacted and scraped, across product groups
  - Less item match across retailers than expected
  - Some items exclusively sold in physical stores, thus no metadata available online



- Working on an additional scraper as long as metadata is not directly received from the retailers themselves

Statistisk sentralbyrå
Statistics Norway

Explicit QA using hedonic regression models

- Four different product groups
  - Mobile phones, laptops, flat screens and computer tablets
  - Hi-tech products, but different degree of technical advancement

- Log-linear specification:

$$\ln p_i^t = a + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k z_{ik}^t + \varepsilon_i^t$$

- Different models tested
  - Aim: Combine high explanatory power combined with practical solutions
    - Possible to reduce the number of variables without reducing too much of explanatory power of the models?

Explicit QA using hedonic regression models II

- Different functional forms: semi-log vs. double-logarithmic
- Categorization of variables vs. continuous variables
  - Risk of loosing information, but at the same time a way of reducing noise in the data set
- Use weighted versions of the hedonic models
  - Different weights provide different results

- Various strategies tested across the different product groups
  - Possible to use only information from scanner data item description text?
    - If computer tablets need only memory and screen size to determine close to 80% variance, are additional variables a necessity?

Price-determining variables

- Automatic model selection vs. more practical knowledge and expert validation

Mobile phones	Laptop computers	Flat screen TVs	Computer tablets
Retailer incl. sales channel	Retailer incl. sales channel	Retailer incl. sales channel	Retailer incl. sales channel
Retailer product group (high-low end)	Retailer product group (type of laptops)	Retailer product group (screen size, inch)	Retailer product group (type of tablets)
Brand	Brand	Brand	Brand
Screen size (inch)	Screen size (inch)	Screen technology	Screen size (inch)
Storage capacity (GB)	Storage capacity (GB)	Smart TV	Storage capacity (GB)
Internal memory (RAM)	Internal memory (RAM)	Resolution	
Display technology	Processor cores	Net weight (KG)	
Number of cameras	Net weight (KG)		

Price index methods

* Short period of data, ongoing project

TDH

Time dummy hedonic method, weighted version

Multilateral method – 13 months of data pooled together

- Index estimate is the coefficient of each time period

DI

Hedonic double imputation method

A more “indirect” method; regression function used to predict prices and incorporated into standard price index formulas

- Combination of matched-model index and the TDH

HP

Homogenous product price indices

Unit values over similar article codes of similar price-determining characteristics

MM

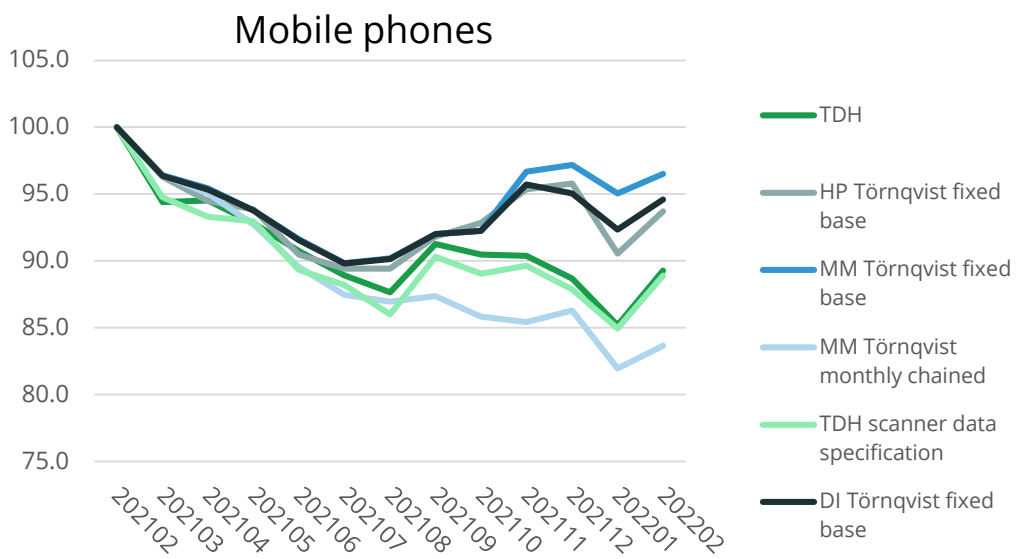
Matched-model price indices

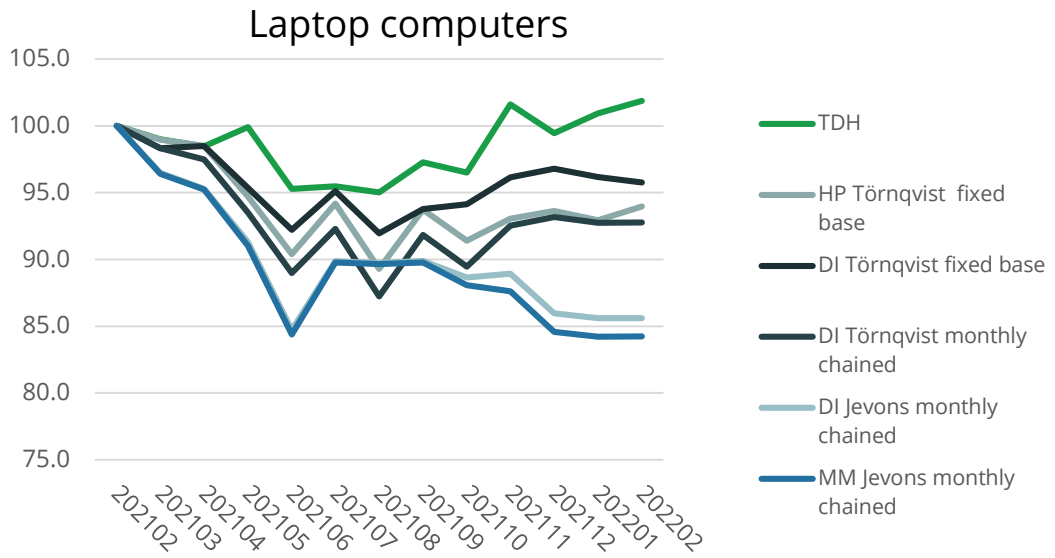
(only) unique article codes are matched over time

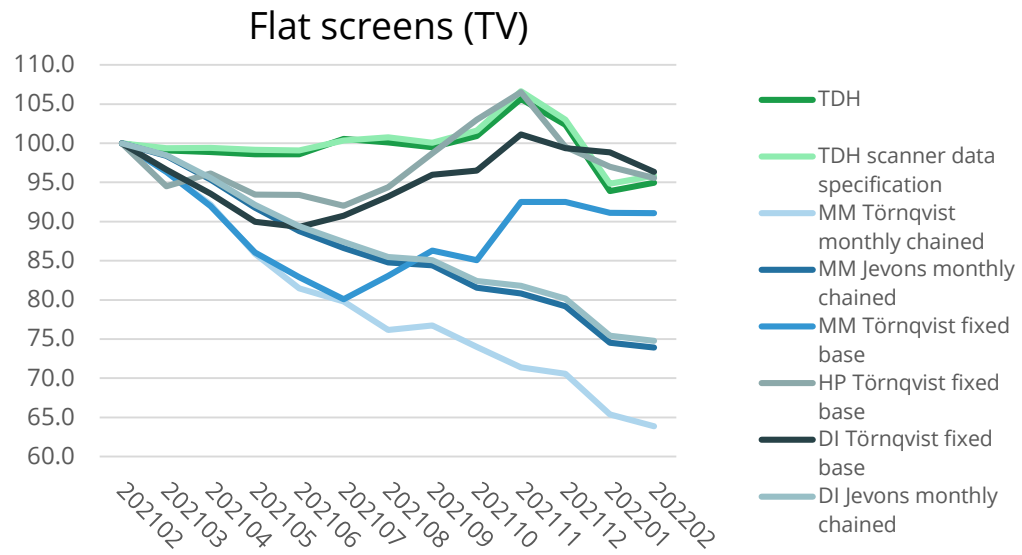
Empirical results

- Different methods provide very different results
  - TDH: all periods included, but fixed effects. Retrospective
  - DI fixed base: increasingly model-based
  - DI m-to-m: minor QA effects, similar to MM m-to-m
  - MM (article code) fixed base: less representativity over time
  - MM (article code) m-to-m: chain drift, as expected
  - HP fixed base: includes new article codes over time, might contain unit value bias

Empirical results II



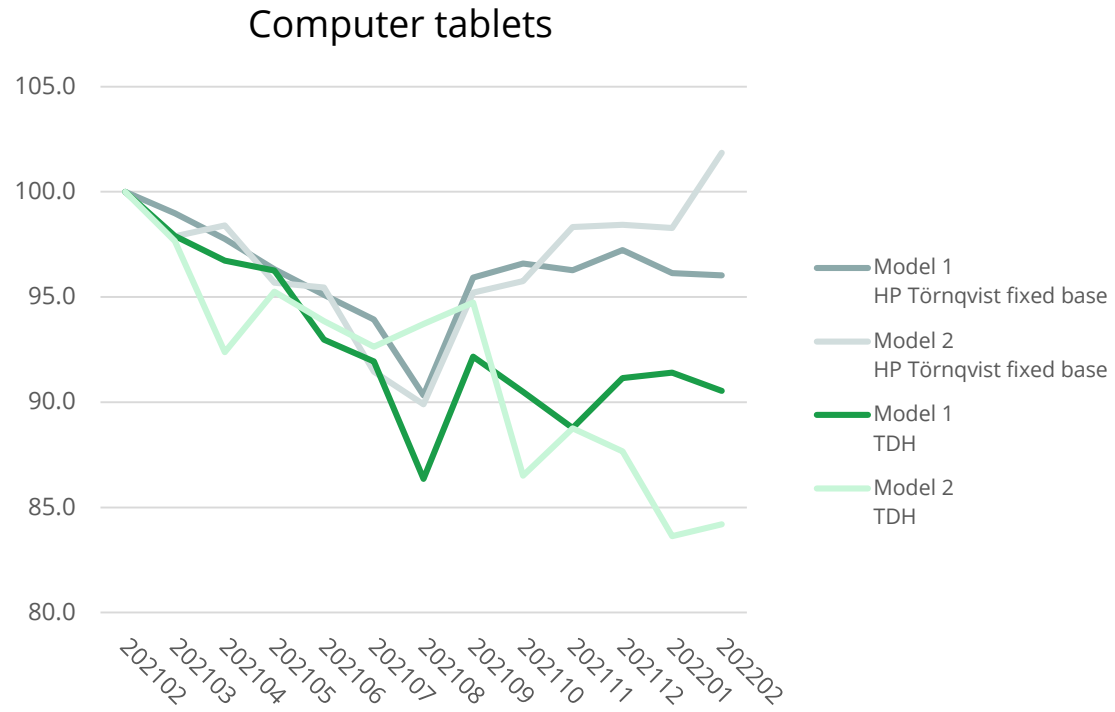




Statistisk sentralbyrå
Statistics Norway

Empirical results III

- Entirely based on metadata extracted from item text in scanner data
- Two variants of model specifications; definition of *brand*
  - Model 1
    - Mother brand and computer tablet family name combined
    - Apple Ipad Pro, Samsung Galaxy tab S4 etc.
  - Model 2
    - Mother brand only
    - Apple, Samsung, Huawei etc.



Conclusions

- Looking for practical solutions for implementing large scale scanner data for consumer electronics
- Combining scanner data and metadata online very resource-intensive
- Study shows promising results for reducing the number of explanatory variables
  - Will reduce time needed for data cleaning, structuring and imputing for missing values
- Consumer electronics not a homogenous product group
  - Implementing hedonics too resource intensive to do for all product categories

→ Likely to use a combination of methods

Statistisk sentralbyrå
Statistics Norway

TALL SOM FORTELLER

Thank you!

More information:
Kjersti.Nyborg.Hov@ssb.no
Ragnhild.Nygaard@ssb.no

Statistisk sentralbyrå
Statistics Norway