



Scannerdata van supermarkten in de consumptieprijsindex

21-03-2024



STATBEL

01

Wat zijn scannerdata?

STATBEL

- Scannerdata
 - Transactiegegevens (incl. kortingen)
 - Supermarkten
 - Gedetailleerde info (barcode)
 - Omzetgegevens en hoeveelheden
- Contract voor gebruik van gegevens door Statbel en voor welke statistieken
- Datasets verschillende retailers niet homogeen
- Ook mogelijk voor consumentenelektronica, kleding, schoenen,...
- Gelijkaardige gegevens kunnen verkregen worden bij marktonderzoeksbureaus

STATBEL

Variabele	Beschrijving	Voorbeeld
DT_STRT	Startdatum week	2/feb/15
DT_STOP	Einddatum week	8/feb/15
CD_PROD_CLASS1	Interne classificatie level 1	D
CD_PROD_CLASS2	Interne classificatie level 2	E
CD_PROD_CLASS3	Interne classificatie level 3	I
CD_PROD_CLASS4	Interne classificatie level 4	K
NR_ITRL	Interne product code - 1	8523
NR_ART	Interne product code - 2	1568
NR_EAN	GTIN code van het product	5449000000286
TX_BRAND_NL	Merk beschrijving - Nederlands	Coca-Cola
TX_BRAND_FR	Merk beschrijving - Frans	Coca-Cola
TX_TYPE_NL	Beschrijving product type - Nederlands	2L
TX_TYPE_FR	Beschrijving product type - Frans	2L
TX_INFO_NL	Beschrijving product info - Nederlands	Regular (PET)
TX_INFO_FR	Beschrijving product info - Frans	Regular (PET)
MS_VAT_RT	BTW-tarief	6
MS_TRNOVR	Omzet	10000
MS_SALES_UNIT	Verkochte hoeveelheid	4000
CD_TYPE	Verkocht per eenheid of per gewicht	Units
MS_PKGG	Waarde verpakking	2
MS_PKGG_DESCR	Beschrijving verpakking (Liter, Stuks, ...)	L
MS_ALC_RT	Alcoholpercentage (%)	0
MS_AV_PRC	MS_TRNOVR/MS_SALES_UNIT	2,5

Variabele	Beschrijving	Voorbeeld
CD_PROD_CLASS1	Interne classificatie level 1	D
TX_PROD_CLASS1	Beschrijving classificatie level 1	Food
CD_PROD_CLASS2	Interne classificatie level 2	E
TX_PROD_CLASS2	Beschrijving classificatie level 2	Drinks
CD_PROD_CLASS3	Interne classificatie level 3	I
TX_PROD_CLASS3	Beschrijving classificatie level 3	Lemonades
CD_PROD_CLASS4	Interne classificatie level 4	K
TX_PROD_CLASS4	Beschrijving classificatie level 4	Regular Cola

STATBEL

- Voordelen scannerdata t.o.v. klassieke prijsopnames
 - Transactiepreizen (incl. kortingen)
 - Omzetinformatie
 - Nieuwe producten onmiddellijk meegenomen in berekening
 - Grotere steekproefgrootte
 - Minder meetfouten
 - Details gewichten lagere niveaus
 - Daling kost prijsopnames

STATBEL

Timeline

- 2012-2013: onderhandelingen met supermarkten
- Okt-2013: start levering wekelijkse gegevens (incl. historische gegevens)
- 2014: Onderzoek
- 2015: Implementatie scannerdata voor aantal productgroepen
- 2016-2017: Uitbreiding aantal productgroepen
- 2018-2019: Onderzoek
- 2020: Toevoeging extra supermarktketen & wijziging methodologie
- 2022: Scannerdata voor consumentenelektronica & huishoudapparaten
- 2024: Uitbreiding aantal productgroepen

STATBEL

02

Methodologie

STATBEL

▪ COICOP-nomenclatuur

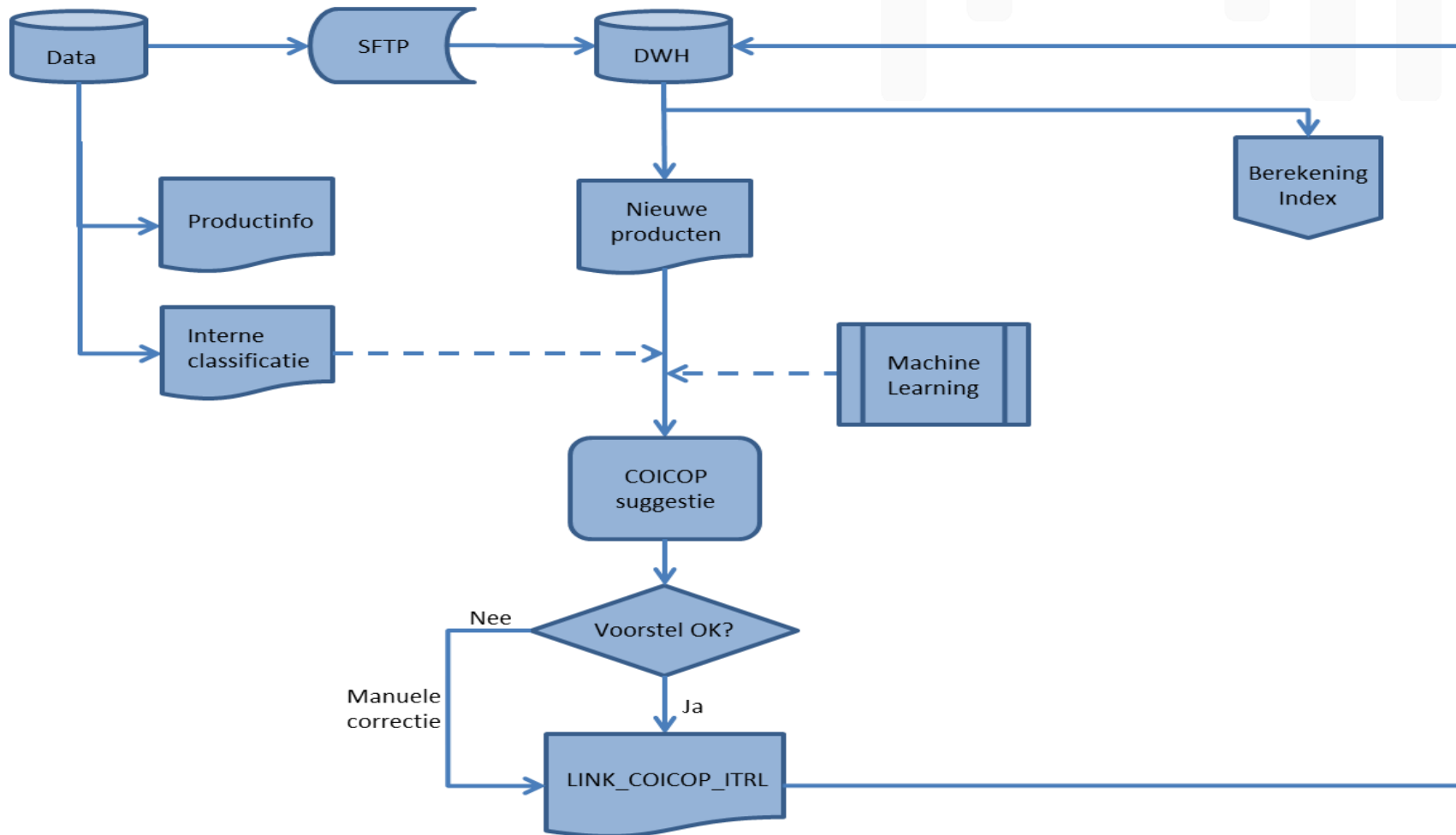
	Hoofdgroepen
1	Voeding & niet-alcoholische dranken
2	Alcoholische dranken & tabak
3	Kleding & schoeisel
4	Huisvesting, water, elektriciteit & gas
5	Meubelen, huishoudelijke app. & onderhoud
6	Gezondheid
7	Vervoer
8	Communicatie
9	Recreatie & cultuur
10	Onderwijs
11	Hotel, restaurant & café
12	Diverse goederen & diensten

COICOP	Benaming	Niveau
0	Totale bestedingen	1
01	Voeding en niet-alcoholische dranken	2
01.1	Voeding	3
01.1.1	Brood en granen	4
01.1.1.1	Rijst	5
01.1.1.2	Meel en andere granen	5
01.1.1.3	Brood	5
01.1.1.4	Andere bakkerijproducten	5
01.1.1.5	Pizza en quiche	5
01.1.1.6	Deegwaren en couscous	5
01.1.1.7	Ontbijtgranen	5
01.1.1.8	Andere producten op basis van granen	5

STATBEL

- Opstartfase:
 - Link interne classificatie met ECOICOP
 - Interne groepen niet altijd bruikbaar -> product linken
- Recurrent: verwerking nieuwe producten
 - Voorstel obv interne classificatie (eventuele correctie)
 - Linking productniveau > linking interne classificatie
- Machine learning (SVM-algoritme)
 - Nieuwe data klasseren obv trainingsdataset

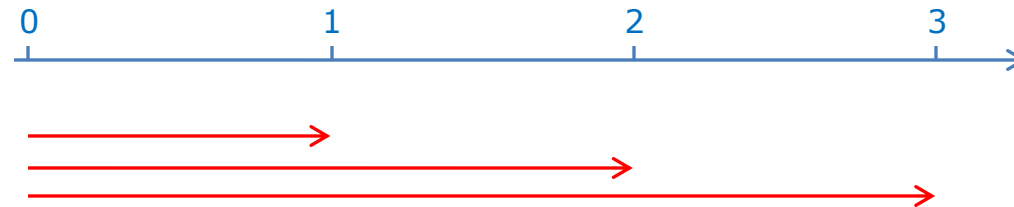
STATBEL



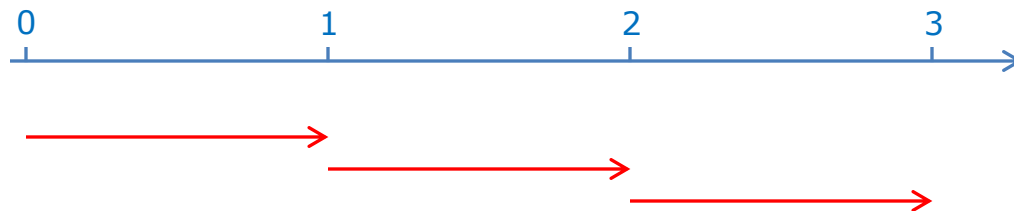
STATBEL

- Twee traditionele manieren om prijsindex te berekenen:

1. Directe index: vergelijking tussen huidige prijs met prijs in basismaand of -jaar

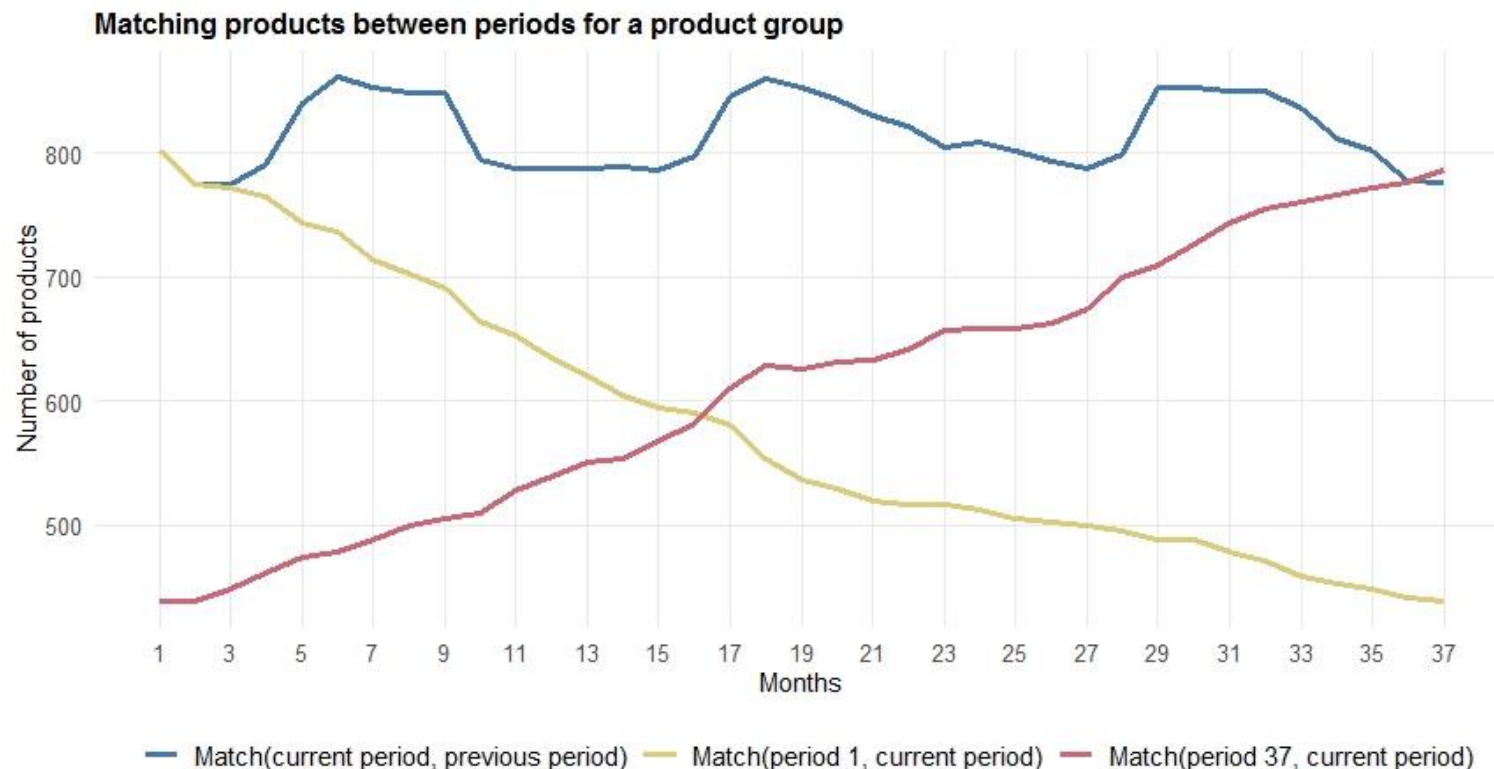


2. Chained index: vergelijking tussen huidige prijs en prijs in vorige maand



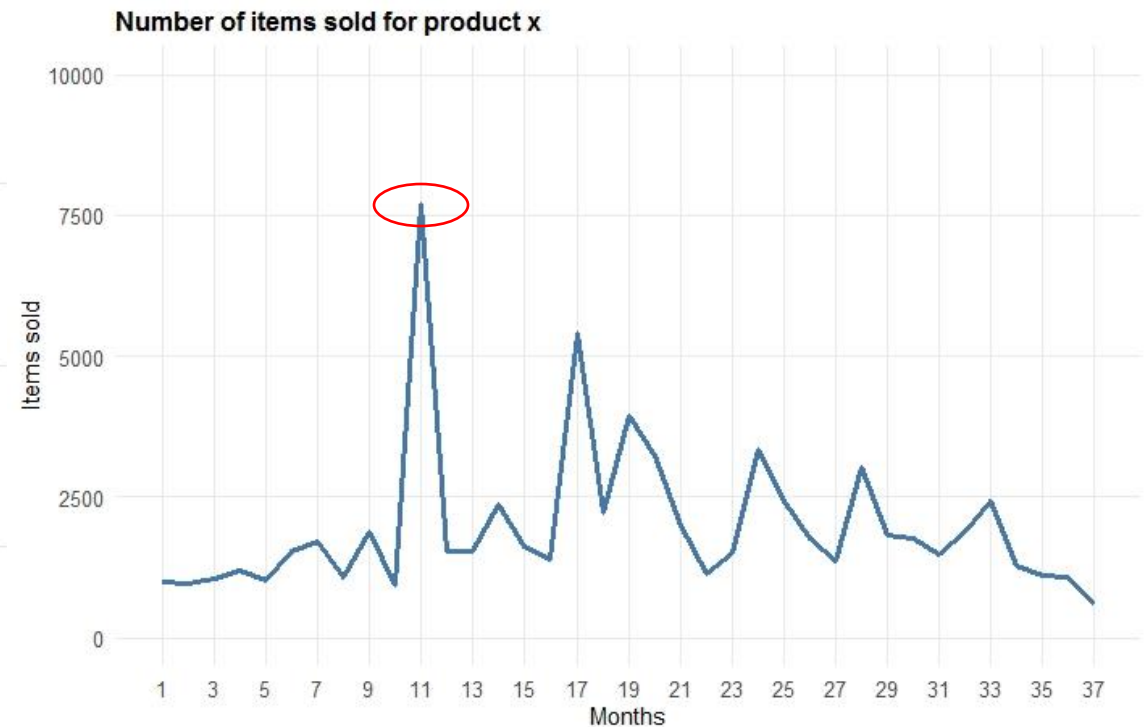
STATBEL

- Scannerdata : hoge 'attrition rate' -> gebruik van directe index onmogelijk (verlies van matching items na verloop van tijd)
- Maandelijks ketting wordt noodzakelijk



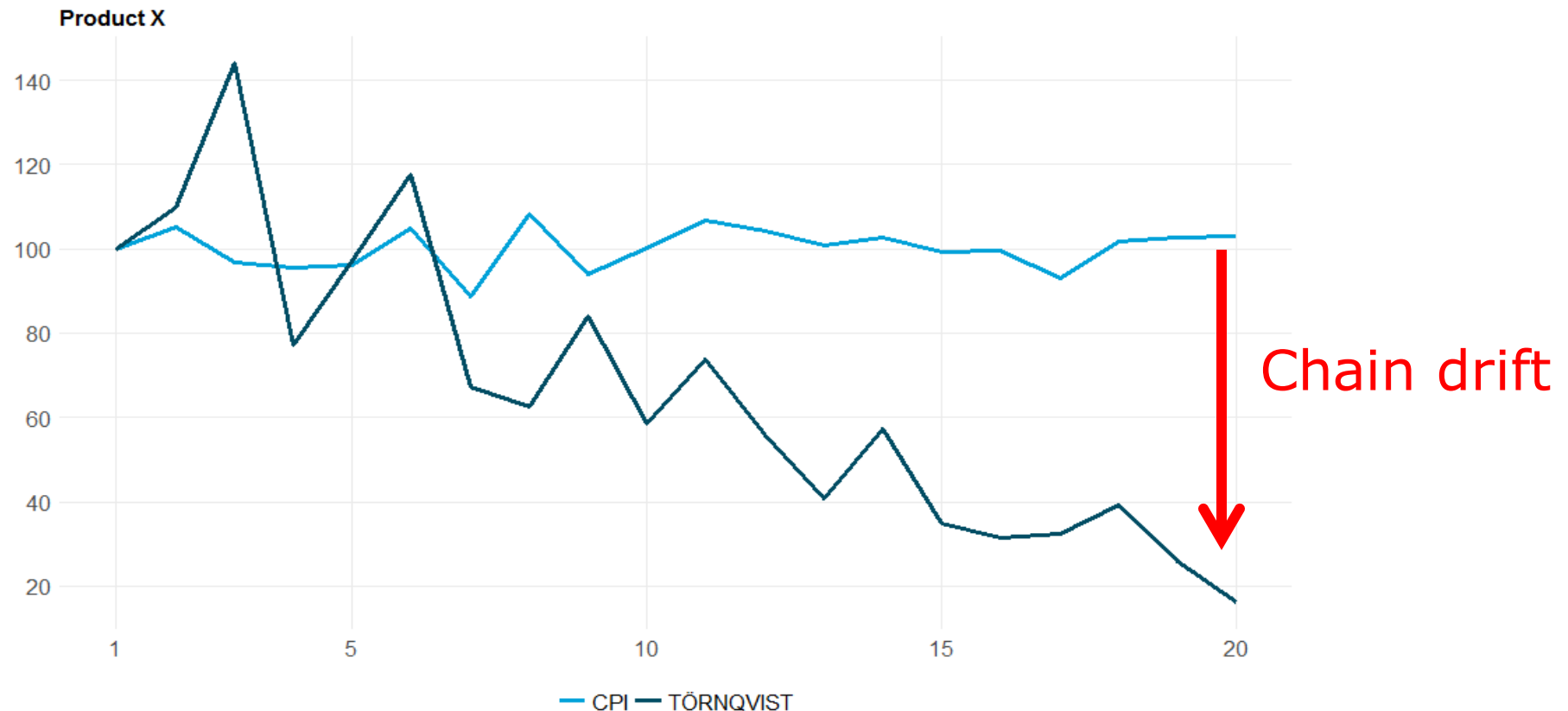
STATBEL

- Prijs- en hoeveelheidsschokken zijn problematisch voor chained indices
- Consumenten reageren op kortingen, maar prijs en hoeveelheid zijn niet uniek bepaald



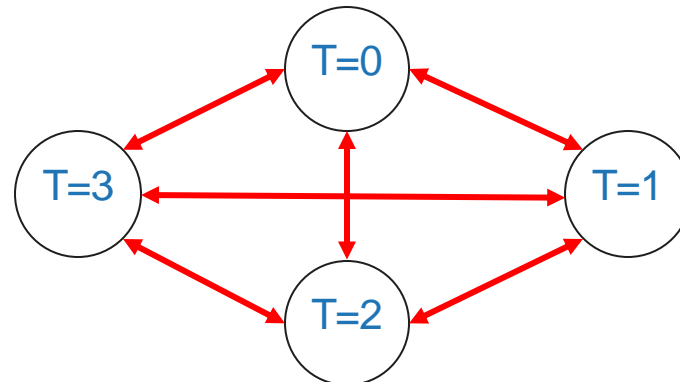
STATBEL

- Chain drift: indices keren niet terug naar 100 als prijs en hoeveelheid identiek zijn tussen 2 periodes → Om dit te vermijden werd initieel een methode gebruikt die niet expliciet rekening hield met omzet



STATBEL

- Ideaal is om toch omzet expliciet te gebruiken
- Onderzoek gebaseerd op spatiale prijsvergelijkingen
 - Soortgelijk probleem: directe prijsvergelijking tussen land A en C moet hetzelfde resultaat geven als indirecte vergelijking van A met C via B
 - Directe index levert hetzelfde resultaat als chaining -> geen chain drift
 - In intertemporele context: $\text{Index}_{1,3} = \text{Index}_{1,2} \times \text{Index}_{2,3}$

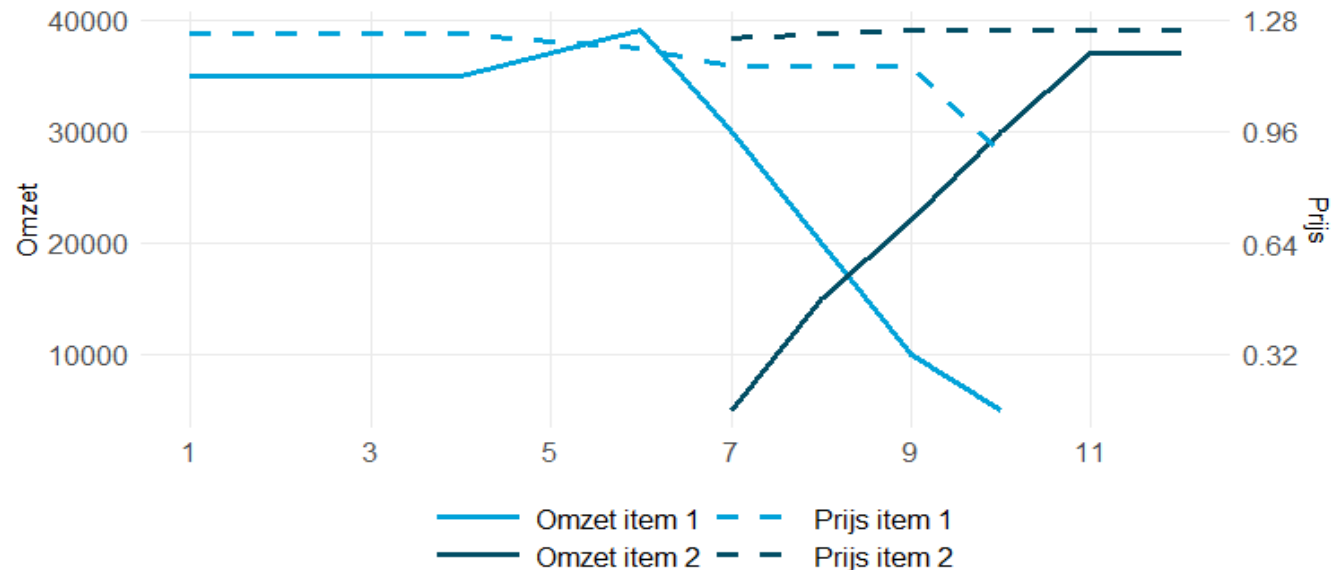


STATBEL

- Aangezien meerdere periodes worden vergeleken => Multilaterale methodes
- Ideale methode?
 - Prijzen wegen op laagste niveau
 - Alle producten meenemen (geen sample)
 - Zo automatisch mogelijk: Quid relaunches?

STATBEL

- Wijziging barcode, maar nog steeds zelfde product
 - Andere verpakking/inhoud
- Ratio prijs/hoeveelheid wijzigt (shrinkflation)
- Link leggen tussen verdwenen en nieuwe producten



STATBEL

03

Internationale context

STATBEL

- Alle landen berekenen consumptieprijnindexen → veel interesse in gebruik van deze gegevens
- Oprichting van expert group bij Eurostat in 2014 (initieel 4 landen, nu een 15-tal landen):
 - 2014-2016: Focus op verkrijgen van gegevens en methodologie
 - 2017-2018: Focus op classificatie, machine learning, ontwikkeling van classificatie applicatie door Eurostat
 - 2019 – heden: Focus op methodologie

STATBEL

- Naast de Eurostat context is er ook de VN
- UN Committee of Experts on Big Data and Data Science for Official Statistics
- Opgericht in 2014
- Mandaat om aanbevelingen te doen bij het gebruik van “big data” voor officiële statistieken

STATBEL

- "UN task team on scanner data" opgericht in 2017 door 6 landen (ondertussen 25-tal landen) → gebruik van deze gegevens wereldwijd doen toenemen
- 3 workstreams:
 - Guidance: ontwikkeling van handboek over indexmethoden
 - Classification: specifiek materiaal over classificatie en machine learning & beschikbaar maken van code
 - Training: cursusmateriaal ter beschikking stellen voor opleidingen over het gebruik van scannerdata voor CPI