

Practices and challenges in using administrative data sources for CPI

CPI WORKSHOP BRUSSELS, OCTOBER 29 - 31



Statistisk sentralbyrå
Statistics Norway

New data sources

Scanner data

- C01 - Food and non-alcoholic beverages
- C02 - Beer, tobacco
- C03 - Clothing
- C05 - Furnishings, household equipment ++
- C06 - Pharmaceutical products
- C07 - Fuels
- C08 - Consumer electronics
- C09 - Recreation goods
- C11 - Restaurants, fast food
- C12 - Personal care products

Web scraping

- ~~C05 - Sewing machines~~
- ~~C07 - Flights~~
- ~~C08 - Consumer electronics~~
- C09 - Recreational goods, electronic games

Adm. sources

- C04 - Electricity
- C07 - Cars
- C07 - Flights, railways and ferries
- C12 - Financial services
- C12 - Social protection



Administrative sources

- **Administrative registers**

- **Road traffic organisation**
 - Newly registered cars (new cars)
- **The Norwegian Directorate for Education and Training**
 - Kindergarten
- **Central registers/Statistics Norway**
 - Business register
- **The Norwegian mapping authorities**
 - Type and number of dwellings
- **The Norwegian tax administrations**
 - National population register

- **Platforms (portals)**

- **The Norwegian Consumer Authority**
 - Electricity
 - Financial services
- **ENTUR** (Government-owned technology company)
 - Railway tickets
 - Bus tickets (trams and metro)
 - Ferries tickets
- **Amadeus** (global travel technology company)
 - Flights



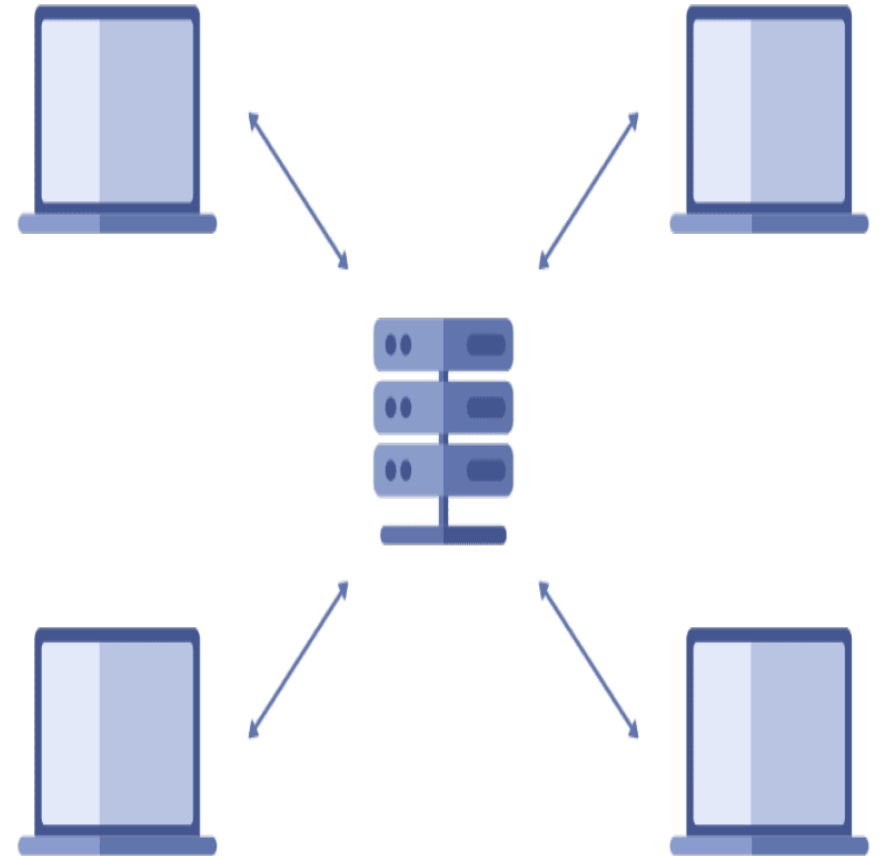
Data collection

- Questionnaires – municipalities
 - Fees services connected to dwellings (Water supply, Refuse- and Sewage collection, etc.)
- Pure Excel files
 - Kindergarten
 - Cars
- Manually from a website
 - Passport fee, medical consultation, physical therapy, education, etc.
- Application Programming Interface (API)
 - Flights, electricity, transport (in pipeline)



Future perspective

- Peer- to-peer
 - Reduce manually data collection
 - Reduce questionnaires



Application Programming Interface (API)

- Taking over for web scraping and “excel files”
 - Flights (longest experience)
 - Railways
 - Electricity (in pipeline)
 - Ferries and busses (testing)

- Portal/platform – contains sample or whole population?



Flights

- New data source and data collection from 2021
- Ready for implementation in January 2021, but postponed due to Covid pandemic and low traveling
- Setting up a new system of data collection during the pandemic – challenging!
- Data from the global booking system **Amadeus**
- Aims:
 - Create stable data collection
 - New calculation method – make use of “big data”



API data set

OSL-TRD	NOK1297.00	TRF (WF/WF484) TRD	2021-06-11T13:25	2021-06-11T14:35	9 K	ECONOMY	FRIDAY	2	06.05.2021	751	KHAPOPS	OCAMP
OSL-TRD	NOK1319.00	OSL (SK/SK332) TRD	2021-06-11T08:05	2021-06-11T09:00	9 O	ECONOMY	FRIDAY	2	06.05.2021	751	OCAMP	KHAPOPS
OSL-TRD	NOK1319.00	OSL (SK/SK380) TRD	2021-06-11T20:40	2021-06-11T21:35	9 O	ECONOMY	FRIDAY	2	06.05.2021	751	OCAMP	KHAPOPS
OSL-TRD	NOK1347.00	OSL (SK/SK338) TRD	2021-06-11T10:50	2021-06-11T11:45	9 T	ECONOMY	FRIDAY	2	06.05.2021	768	THAPPY	THAPOP
OSL-TRD	NOK1347.00	OSL (SK/SK344) TRD	2021-06-11T13:20	2021-06-11T14:15	9 T	ECONOMY	FRIDAY	2	06.05.2021	768	THAPPY	THAPOP
OSL-TRD	NOK1347.00	OSL (SK/SK362) TRD	2021-06-11T15:35	2021-06-11T16:30	9 T	ECONOMY	FRIDAY	2	06.05.2021	768	THAPPY	THAPOP
OSL-TRD	NOK1347.00	OSL (SK/SK364) TRD	2021-06-11T17:00	2021-06-11T17:55	9 T	ECONOMY	FRIDAY	2	06.05.2021	768	THAPPY	THAPOP
OSL-TRD	NOK1347.00	OSL (WF/WF779) TRD	2021-06-11T19:55	2021-06-11T21:05	2 T	ECONOMY	FRIDAY	2	06.05.2021	768	THAPOP	THAPPY
OSL-TRD	NOK1347.00	OSL (WF/WF779) TRD	2021-06-11T19:55	2021-06-11T21:05	2 T	ECONOMY	FRIDAY	2	06.05.2021	768	THAPOP	THAPPY



Advantages Amadeus API

- Old methods: web scraping from each flight company
- More stability in using API compared to web scraping web sites
 - Less maintenance work
- Increased coverage – more flight companies compared to before
- Increased efficiency – only one single data set and only one data supplier
- *Bulk scraping* – scrape as much relevant information as possible



Information downloaded

- Weekly data files from Amadeus API «*Flight Offers search*»
- Available tickets – real time prices
 - «What is the cheapest flight from Oslo to Madrid 1 of July 2021?»
- Pre-defined searches or trips based on information from the airfare industry
- Important to reduce the searches/trips to have an efficient and manageable data set

The Amadeus logo, consisting of the word "amadeus" in a white, lowercase, sans-serif font, centered on a solid blue rectangular background.The Amadeus for Developers logo, featuring the word "amadeus" in blue and "for Developers" in a smaller, lighter blue font, both in a sans-serif typeface.

Statistisk sentralbyrå
Statistics Norway

Information downloaded, cont.

- Round trips
- Economy tickets
- Direct flights
- Representative trips defined within Norway (from Oslo to the largest cities) and from Oslo to different cities abroad
- Information on price, flight nr, company, airport, departure- and arrival time, ticket type, booking category, time of data collection, luggage



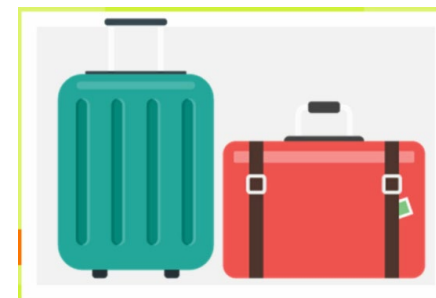
Time of ordering and time of travelling

- Flights included in the CPI when consumed, not purchased
- Separate between time of ordering and time of travelling
 - Especially when prices change frequently
- Missing information on when private consumers book their flights
 - Especially during a pandemic/post-pandemic!
- Data collection starts 3 month ahead for domestic flights
- Time of travelling/departure for all trips in relation to weekend



Price concept

- Cover several weeks of the months
 - Volatile prices on tickets
- Consumer prices including taxes
 - Actual price paid by consumer
- Include additional taxes that are compulsory
- Measure changes in luggage fees
- Some deviations in API prices and prices on companies own web pages
 - API does not cover 100% of all available tickets (not all sold tickets available in Amadeus)
 - Still relatively good correlation



Calculation method

- Average price based on different flights with the same characteristics
- Flights stratified into time intervals (morning, mid-day, afternoon, night)
 - Estimate average prices within these intervals
- Assuming consumers are indifferent across companies at the same destination/location
 - Assuming consumers are looking for the cheapest ticket independently of the airline company
 - Strong assumption



Average prices

Destination	Time of booking	Time of day	202100	202103
OSL-TRD	1	1	1314,47975	1473,40655
OSL-TRD	1	2	1517,90008	1464,01417
OSL-TRD	1	3	1441,05954	1515,7596
OSL-TRD	1	4	842,002987	878,86
OSL-TRD	2	1	1273,64573	1228,35646
OSL-TRD	2	2	1347,58214	1281,17425
OSL-TRD	2	3	1311,5404	1233,18397
OSL-TRD	2	4	1318,75709	1083,66667



Further aggregation of prices

- Prices aggregated to different destinations across time of booking and time of day
 - Destinations and weights from AVINOR surveys (government company responsible for the domestic airports)
- Lacking information about when private consumers make their booking
 - assume that bookings are made evenly throughout the booking months ahead
- Lacking information about when consumers travel during the day
 - Estimate unweighted average of the different time intervals



Sub price index formula

- Price index of domestic flights $P^{t/0}$;

P_i^t average price for a certain domestic destination for a certain period, P_i^0 average price the same destination in the price reference period, w_i^b weight for destination:

$$P^{t/0}_{\text{elementary index}} = \sum w_i^b * \frac{P_i^t}{P_i^0} \text{ where } \sum w_i^b = 1$$



Index formula, cont.

- Destinations and corresponding weights based on non-business travellers
 - Data from AVINOR (government company responsible for the domestic airports)
- Sub price index for domestic and international flights weighted together to index of airline fares
 - Not published separately - confidentiality reasons



International flights – exceptions

- Calculation as the domestic sub index with some exceptions
- Data collection starts 7 month ahead for international flights
- Flights not stratified into detailed time intervals
 - Fewer options on non-Nordic flights
 - Geometric average of all available ticket prices within the day of traveling



Questions?

