Statistics Poland

Voorburg
Copenhagen 2025

Use of non-public data sources for compiling SPPI for ISIC 4922 Other passenger land transport – taxi operation

Industry paper

The 39th Voorburg Group Meeting
22nd -26th September, 2025

dr Agnieszka Matulska-Bachura
Beata Cebula

Co-funded by the European Union

Plan of presentation

1. Introduction
2. Description and characteristics of industry
3. Measurement of SPPI
4. Methodology for calculating taxi services price indices based on web-scraped data
5. Conclusions

Introduction

- Objective – presenting the practices&experiences of the Statistics Poland in using the web-scraped data for compiling SPPI for taxi operation being the part of part of ISIC 4922 *Other passenger land transport* (by NACE Rev.2 4932 *Taxi operators*);
- Focus on the traditional taxi services;
- Constraints in monitoring the ride-hailing market;

Description and characteristics of the industry - classifications

Taxi operation

ISIC REV.4	NACE REV.2
Section H Transportation and storage	Section H Transportation and storage
Division 49 Land transport and transport via pipelines	Division 49 Land transport and transport via pipelines
492 Other land transport	493 Other land transport
4922 Other passenger land transport	4932 Taxi operation
▪ *scheduled long-distance bus services,* ▪ *charters, excursions and other occasional coach services,* ▪ ***taxi operation,*** ▪ *airport shuttles,* ▪ *other renting of private cars with drivers,* ▪ *operation of school buses and buses for transport of employees, etc.*	▪ ***taxi operation,*** ▪ *other renting of private cars with driver;*

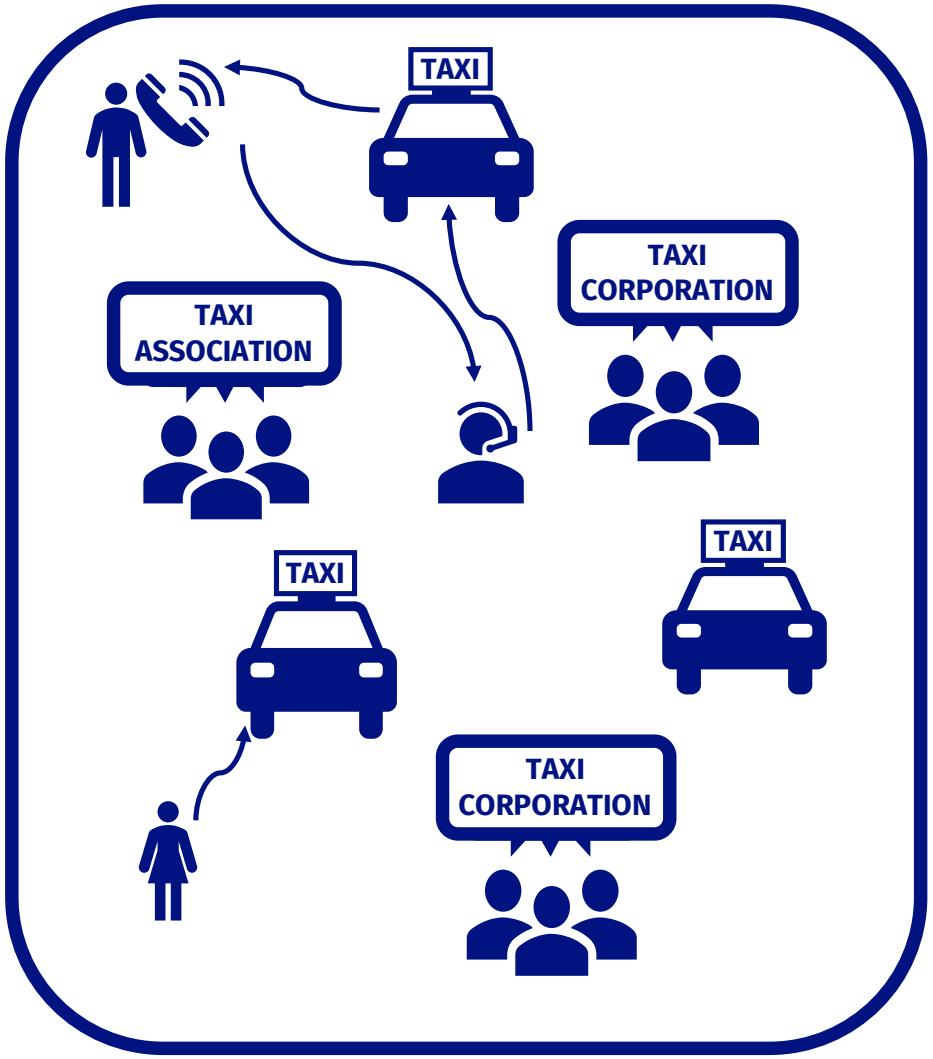
Co-funded by the European Union

Description and characteristics of the industry - classifications

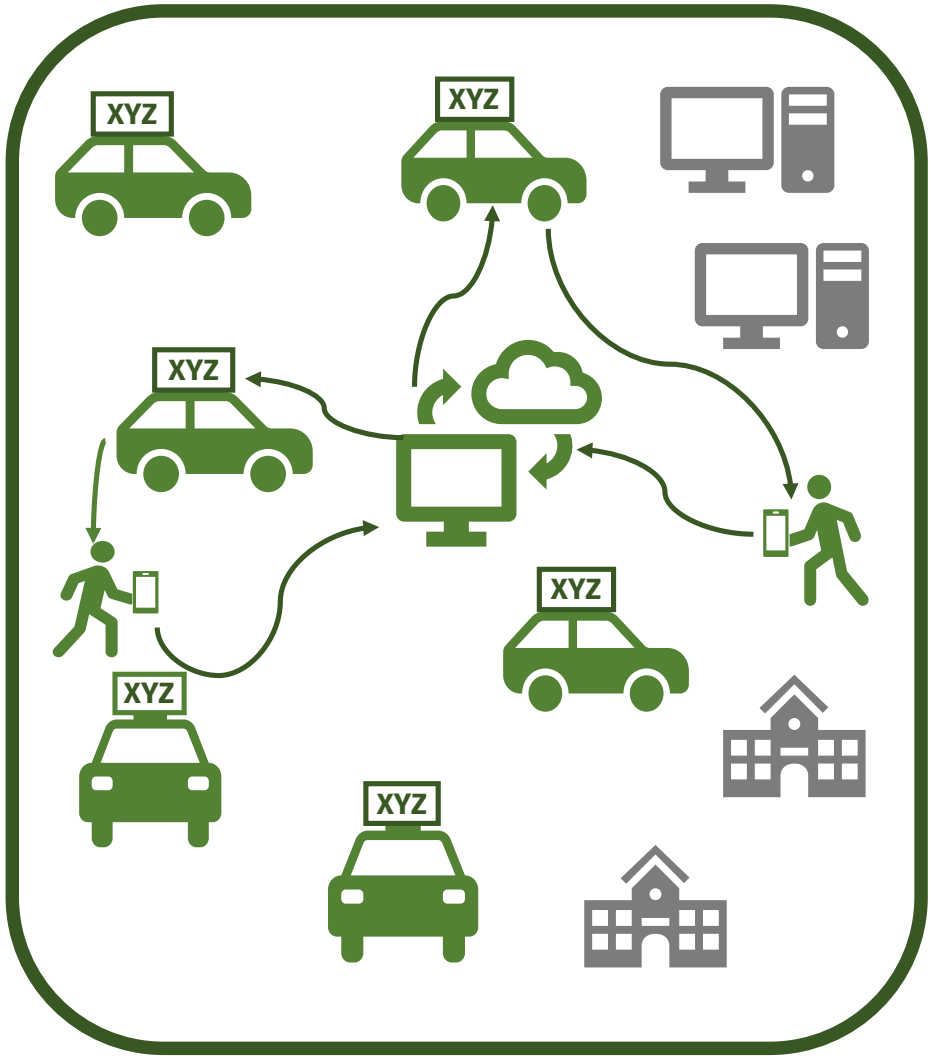
Taxi services

CPC	CPA
Section 6 Distributive trade services; accommodation, food and beverage serving services; transport services; and electricity, gas and water distribution services	Section H Transportation and storage services
Division 61 Passenger transport services	Division 49 Land transport services and transport services via pipelines
641 Local transport and sightseeing transportation services of passengers	493 Other land transport services
6411 Urban and suburban land transport services of passengers	4932 Taxi operation services
▪ … ▪ 64115 Taxi services, ▪ 64116 Rental services of passenger cars with operator, ▪ ….	▪ 49.32.1 Taxi operation services: ✓ 49.32.11 Taxi services, ✓ 49.32.12 Rental services of passenger cars with driver;

Description and characteristics of the industry



TRADITIONAL TAXI SERVICES

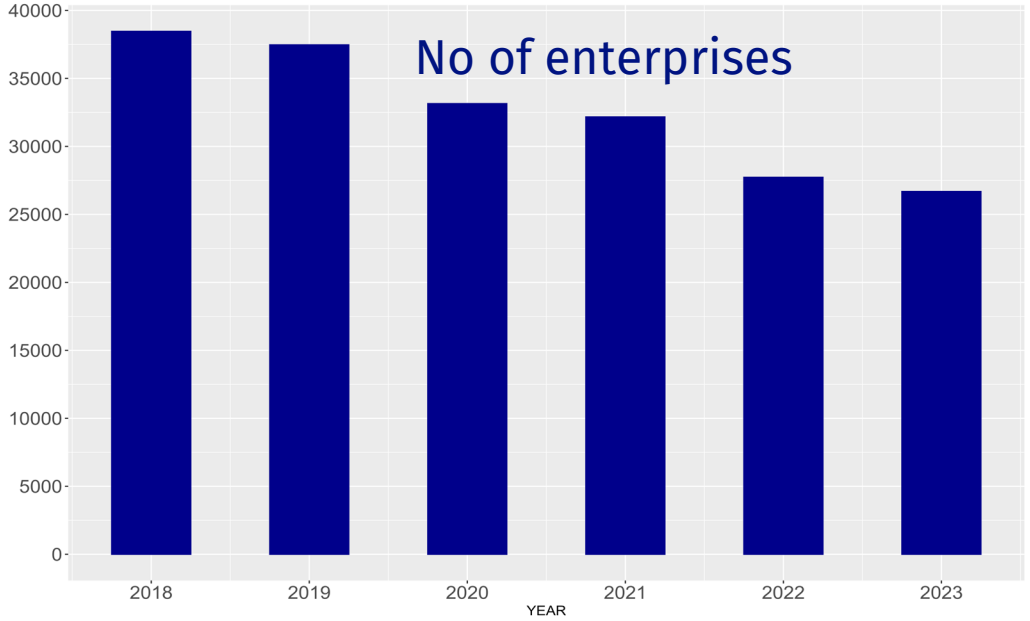


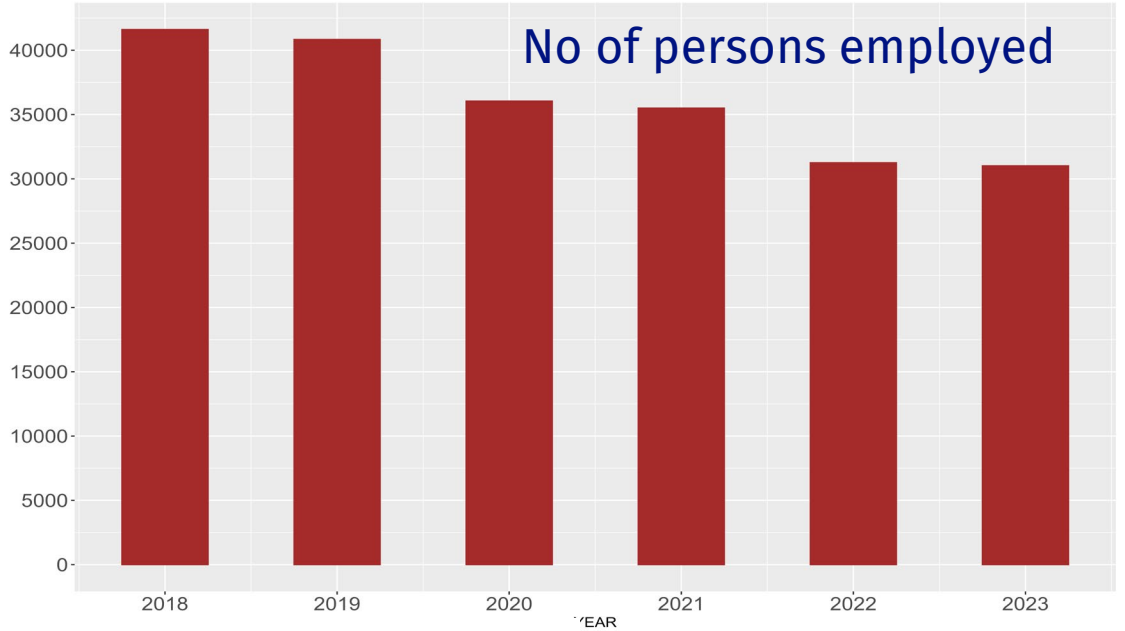
RIDE-HAILING SERVICES

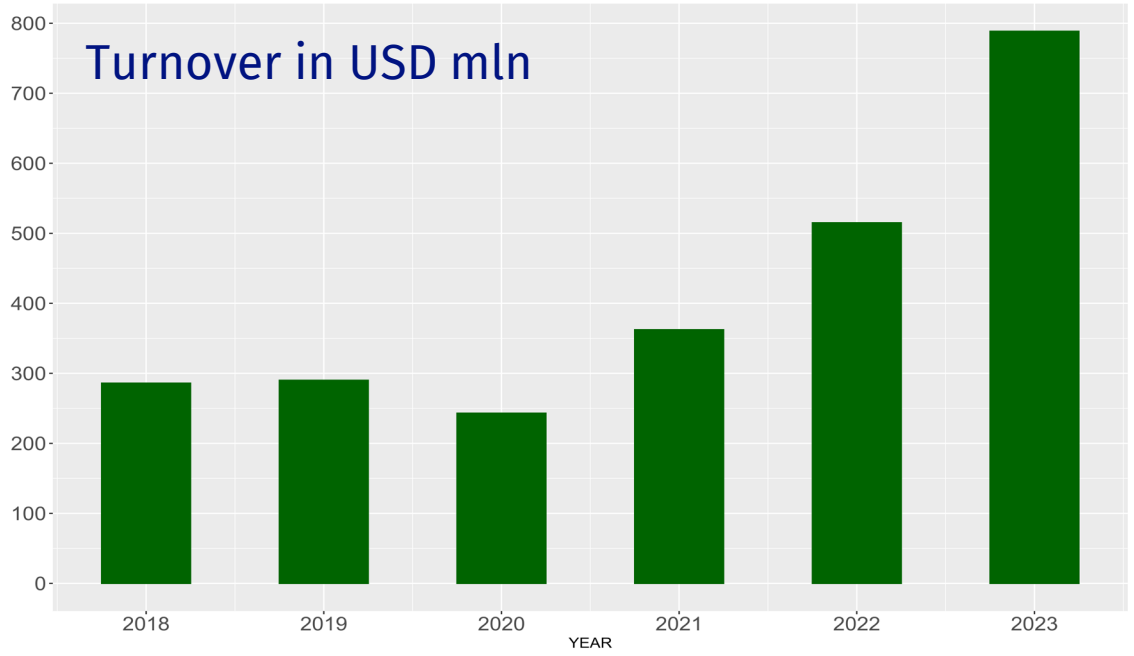
Statistics Poland

Co-funded by the European Union

Description and characteristics of the industry







Statistics Poland

Co-funded by the European Union

Measurement of SPPI

- General framework:
  - one of the obligatory variables compiled within the European Statistical System in compliance with the EBS Regulation;
  - produced quarterly for division 49 *Land transport and transportvia pipelines;*
  - transmitted to Eurostat within 90 days after ending the reference quarter;
  - applied as a deflator for various nominal variables in currnet prices, for example: turnover, national accounts;
- Meausrement issues:
  - SPPI for passenger land transport (division 49) compiled as a part of regular survey of official statistics - *Survey on services producer prices* (since 1999);
  - covers units with no of employees of 10 persons or more, conducting both primary and secondary economic activities classified according to NACE Rev.2 into classes:
    - ✓ 49.10 Passenger rail transport, interurban,
    - ✓ 49.31 Urban and suburban passenger land transport,
    - ✓ **49.32 Taxi operation,**
    - ✓ 49.39 Other passenger land transport not elsewhere classified;

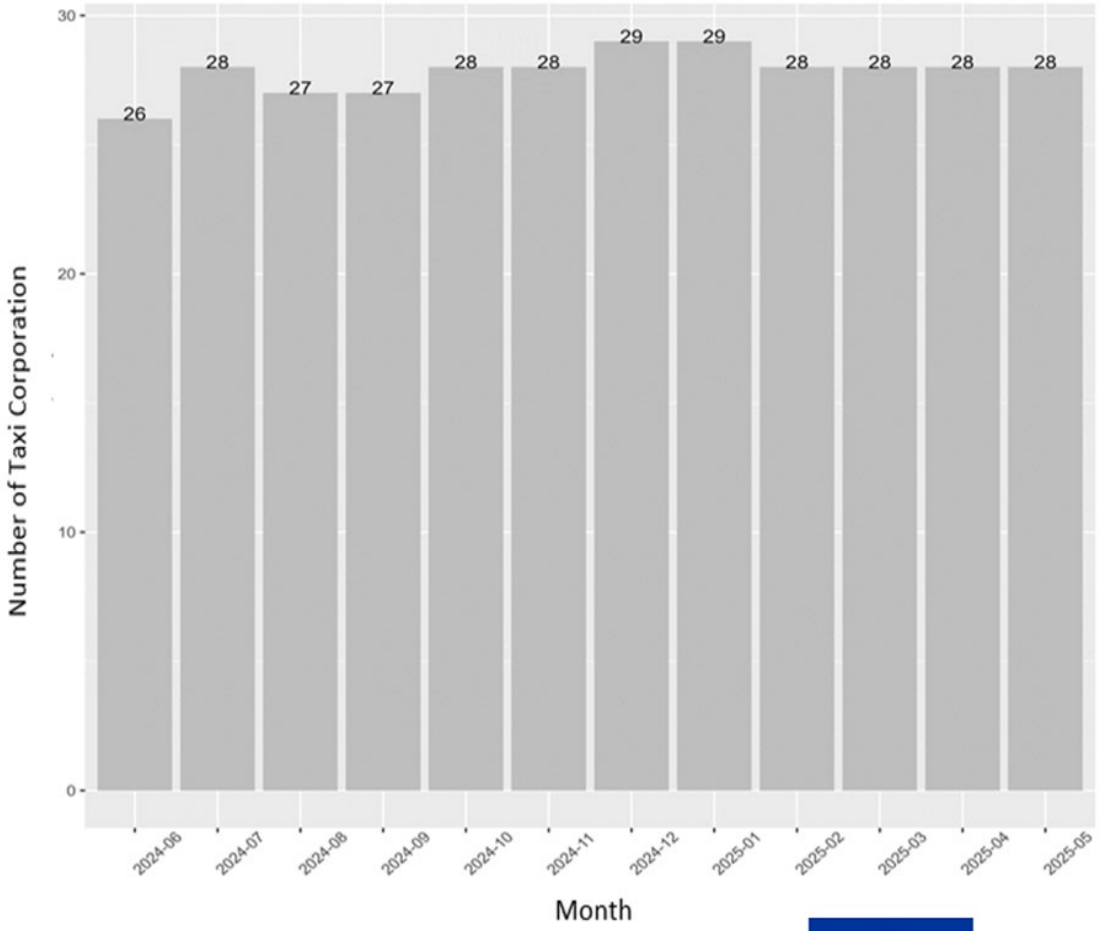
Measurement of SPPI

- Measurement issues (cont.):
  - Data sources:
    - ✓ Electronic on-line application filled in by respondents (C09 *Report on producer prices of transport, storage and telecommunications services*) - for 49.31 *Urban and suburban passenger land transport*&49.39 *Other passenger land transport not elsewhere classified*;
    - ✓ Price lists - for 49.10 *Passenger rail transport, interurban*&**49.32 *Taxi operation*;**
- Development works:
  - undertaken in 2024 by the Statistics Poland in order to improve the process of data collection and processing (under the agreement on TIMELIER, MORE RELEVANT AND MORE INTEGRATED EUROPEAN BUSINESS STATISTICS in the area of The *Improvement of European business statistics production processes);*
  - after review of taxi corporations' websites 35 of them were preliminary selected for web-scraping (finally 29 websites are web-scraped);
  - information collected on daily basis: city, name of the company, email address of the price list, date of listing, initial charge and price per km in tariff 1;
  - tool for automated data collection (for web-scrapping) developed as well as database for data warehousing created;
  - period: June 2024 – May 2025;

Measurement of SPPI

Cities with the web-scraped websites of taxi operators in the period June 2024 - May 2025

No of web-scraped taxi operators in the period June 2024 - May 2025



Co-funded by the European Union

Methodology for calculating taxi services price indices based on web-scraped data

Three key elements:

<u>determining the representative service;</u>

selecting the approach to constructing the index;
establishing a system for weighting price indices for cities at the national level;

10 km ride under tariff 1

price = initial charge + 10 (km) x price for 1 km ride under tariff 1
(reduced by VAT)

Statistics Poland

Co-funded by the European Union

Methodology for calculating taxi services price indices based on web-scraped data

Three key elements:

- ❑ determining the representative service;
- ❑ **selecting the approach to constructing the index;**

establishing a system for weighting price indices for cities at the national level;

- monthly price of the service calculated as the geometric mean of daily quotations;
- three approaches to the index formula considered:

① **the so-called unit value index** – when calculating the index all observations in the period under review (numerator) and all observations in the base period (denominator) are taken into account;

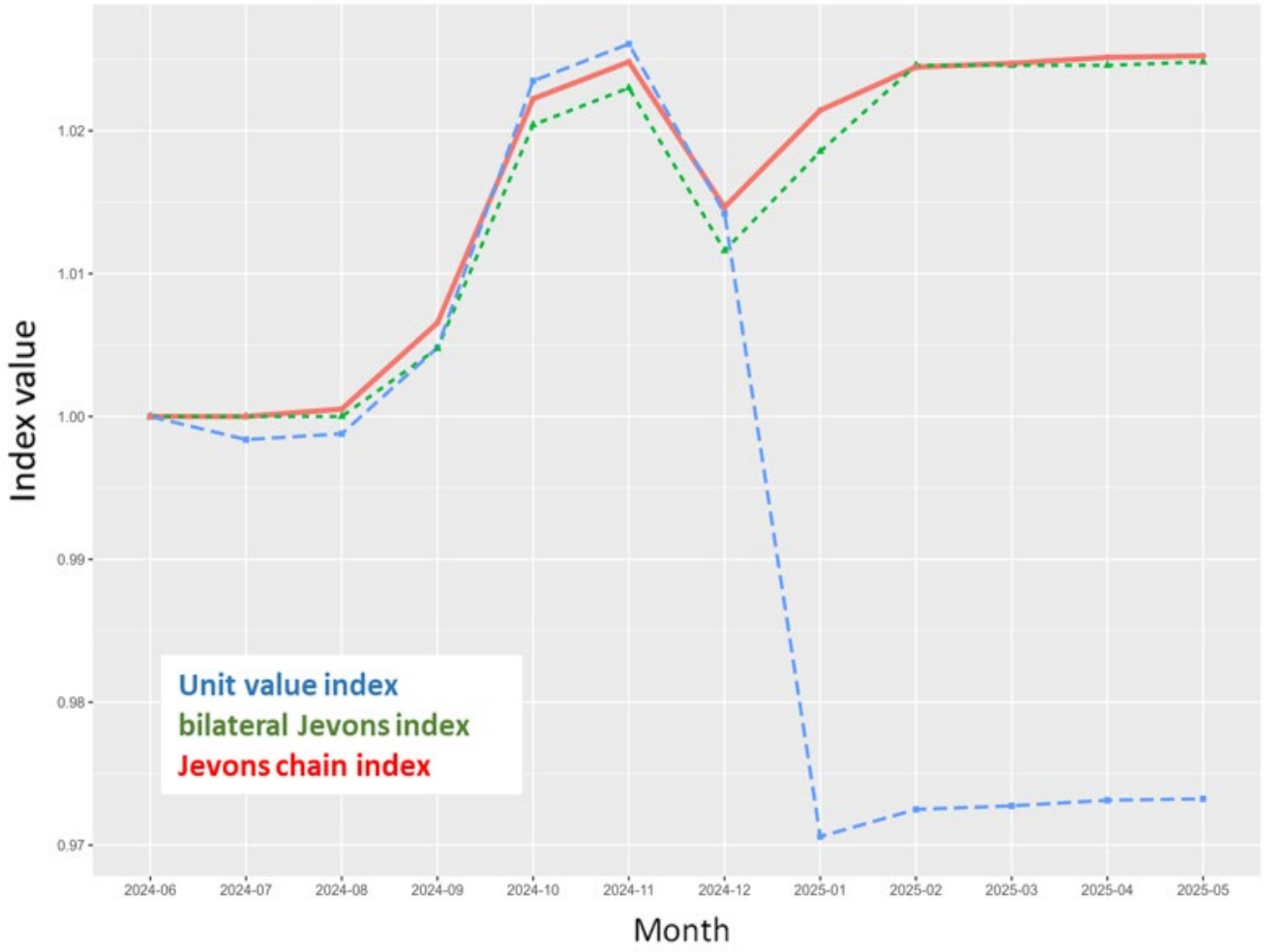
② **the bilateral Jevons index** – the index is calculated on the basis of observations occurring simultaneously in the period under review and the base period;

③ **the chain Jevons index** – for each period under review, price change indices are calculated in relation to the previous period for common observations, i.e., within a given city for the same corporations, and then the indices in relation to the base period are calculated using the chain method;

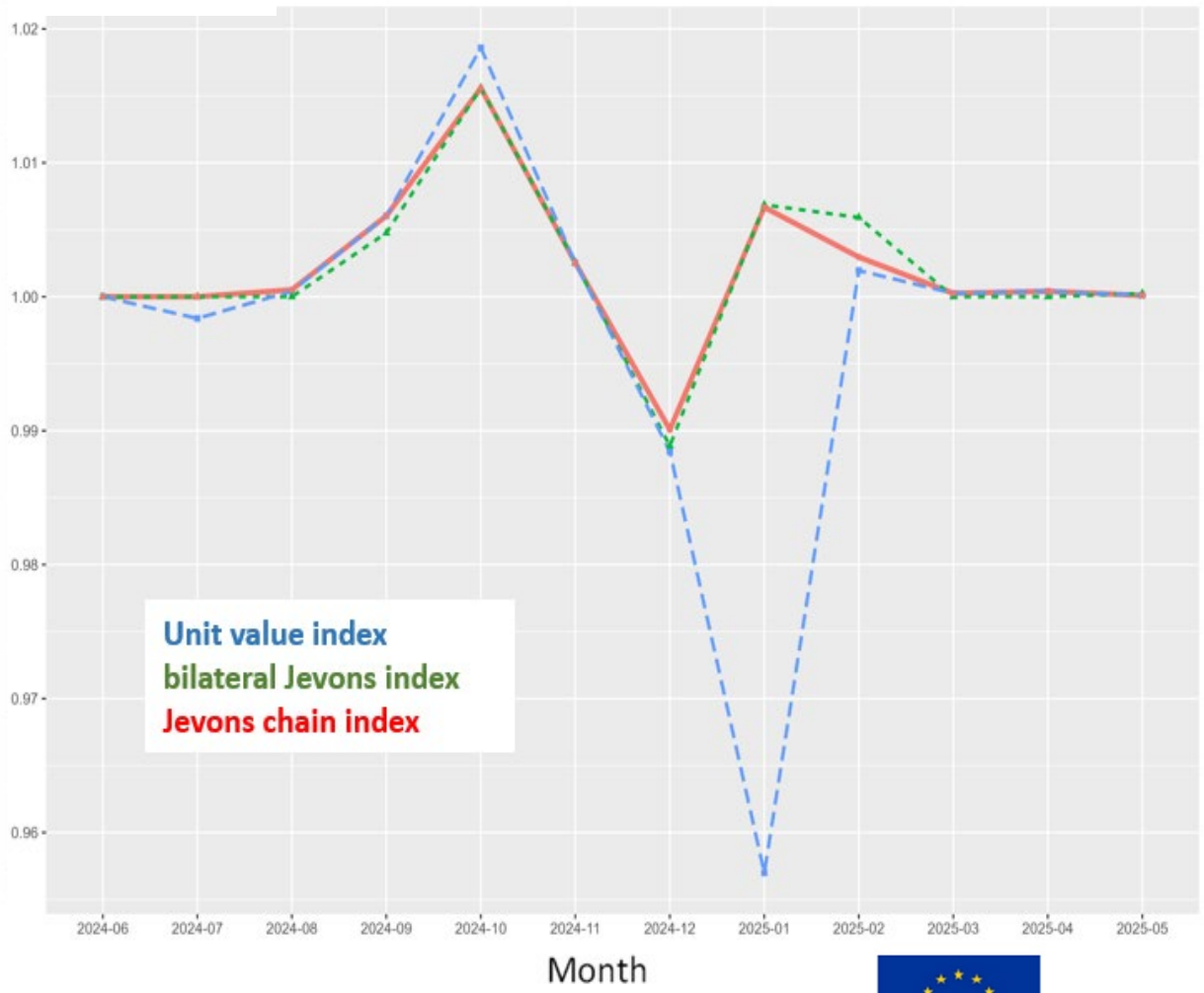
Methodology for calculating taxi services price indices based on web-scraped data

The experimental price indices for taxi operators activity by type of index formula

M0=100



M-1=100



Statistics Poland

Co-funded by the European Union

Methodology for calculating taxi services price indices based on web-scraped data

Three key elements:

- ❑ determining the representative service;
- ❑ selecting the approach to constructing the index;
- ❑ **establishing a system for weighting price indices for cities at the national level;**

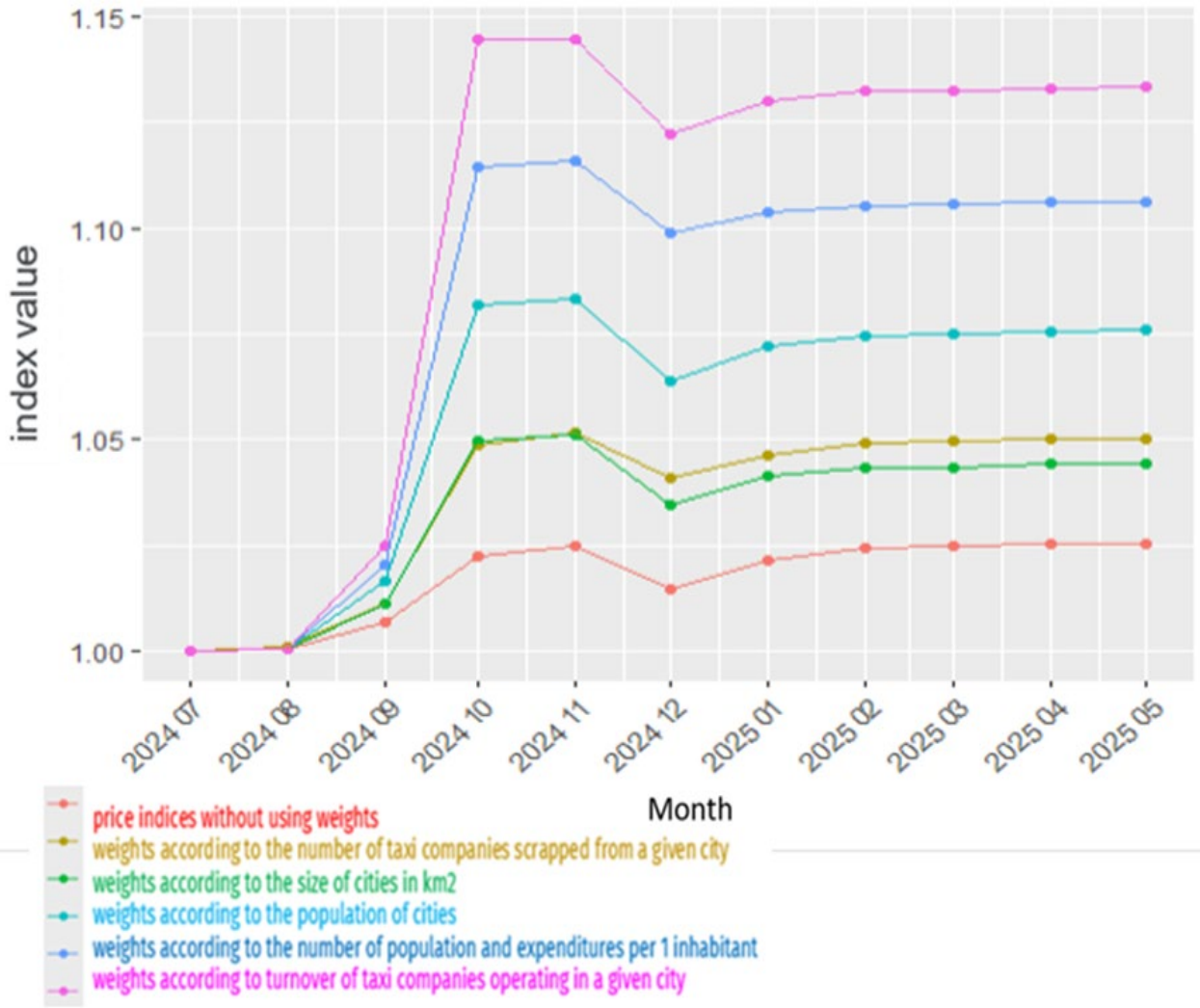
- Two stages procedure for calculating price indice:
  - ✓ 1. the price for the city calculated for each month as the geometric mean of the prices quoted for corporations,
  - ✓ 2. Then, the price for the country compiled as the simple or weighted geometric mean of prices for cities;
- The considered options for determining weights:
  - ✓ according to the number of taxi companies registered in a given city,
  - ✓ according to the population of cities,
  - ✓ according to the size of cities measured in km 2,
  - ✓ by population and per capita expenditure,
  - ✓ by turnover of taxi companies operating in a given city,
  - ✓ **no weights – unweighted price index – geometric mean of prices for cities;**

Statistics

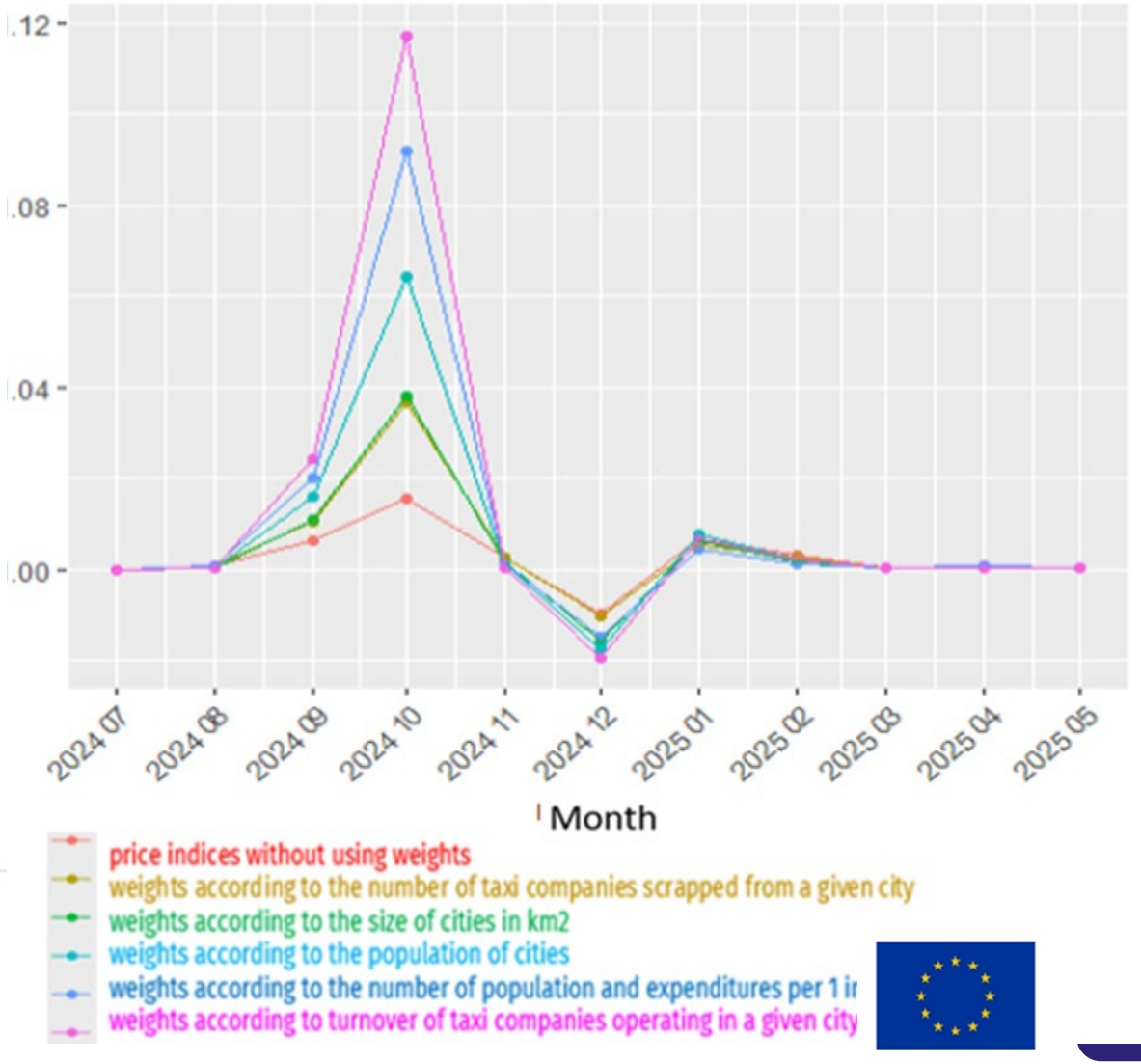
Methodology for calculating taxi services price indices based on web-scraped data

The experimental price indices for taxi operators activity by type of weighting scheme

M0=100



M-1=100



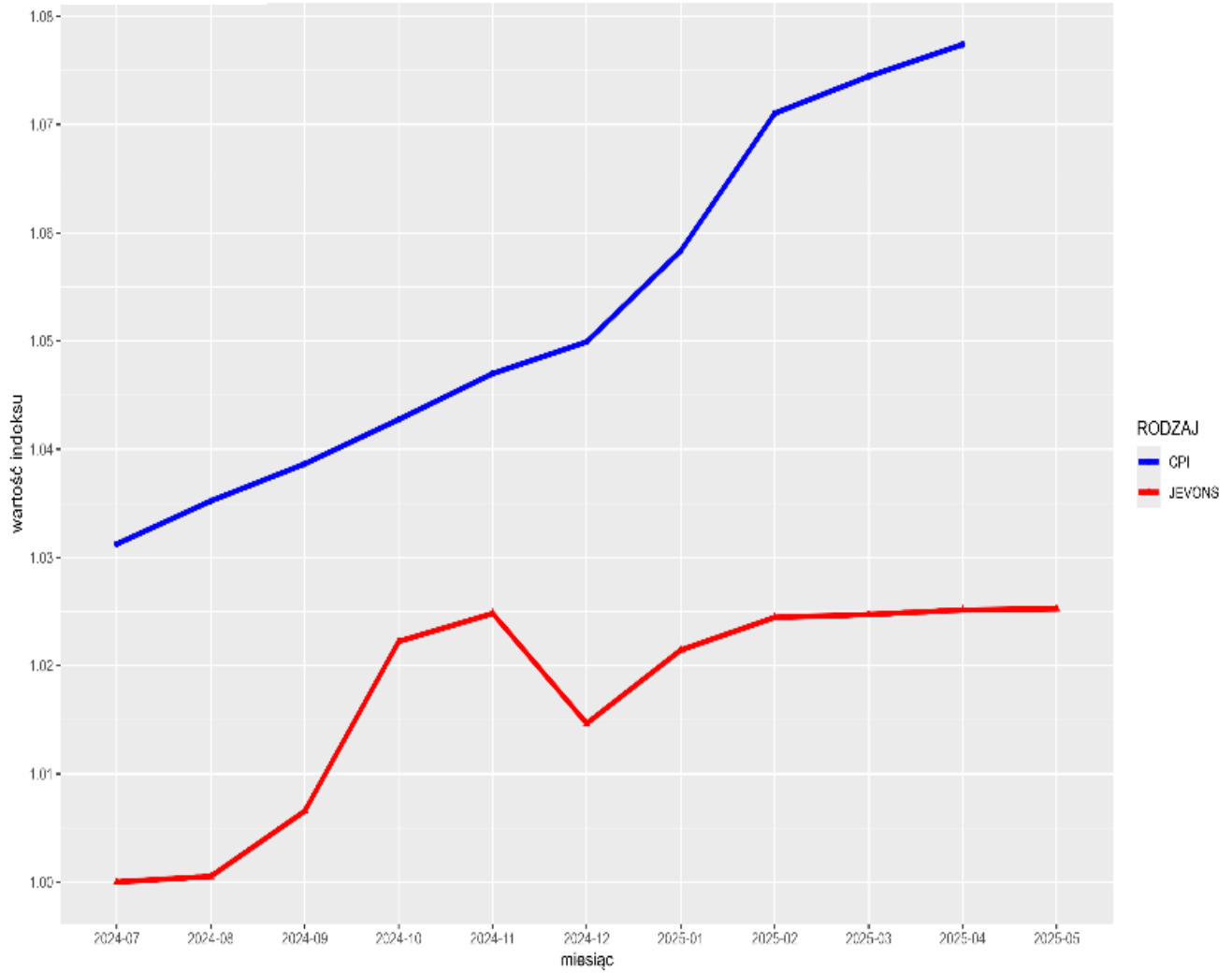
Co-funded by the European Union

Methodology for calculating taxi services price indices based on web-scrapped data

The experimental price indices for taxi operators activity by type of weighting scheme

M-1=100

M0=100



Conclusions:

1) Use of web-scraped data in price statistics recommended, however, many efforts needed when implemented into the regular statistical production;
2) Data from web-scraping as a valuable sources of data for official statistics: offering access to a greater amount of information ~ to a larger number of representatives, more efficient & faster way for data collection than traditional statistical reporting;
3) However, the implementation of new data sources poses challenges for official statistics:
   - developing a tool for automatic data collection & ongoing monitoring of process,
   - creating a database and ensuring the sufficient server space for storing&processing data;
   - updating the tool and database;
4) Necessity for adapting the methodology of price indices calculations to the specifics of web-scraped data - methodological solutions used for traditional data sources not applicable;
5) The implementation of IT tools for data processing &price indices calculation (for example R) required;

Statistics Poland

Thank you for your attention

dr Agnieszka Matulska-Bachura
Trade and Services Department
A.Matulska@stat.gov.pl

Beata Cebula
Trade and Services Department
B.Cebula@stat.gov.pl

Co-funded by the European Union