Statistics Poland

Trade and Services Department

Industry paper

The 39th Voorburg Group Meeting
Copenhagen

22th-26th September 2025

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

by
Agnieszka Matulska-Bachura
Beata Cebula

Statistics Poland
Trade and Services Department

Aleja Niepodległości 208, 00-925 Warsaw
phone +48 22 608 31 24
Sekretariat-HU@stat.gov.pl
stat.gov.pl

Statistics Poland

Table of content:

The main purpose of paper is to present practices and experiences of the Statistics Poland in using of web-scrapped data for compiling SPPI for taxi operation as a part of ISIC 4922 *Other passenger land transport* (by NACE Rev.2 4932 *Taxi operators*).

Due to the constraints resulted from the characteristics of taxi operation market in Poland the works have concentrated on the market of traditional taxi services. Currently, it is not possible to cover the majority of ride-hailing market with observations. The companies operating in that segment of market are not residents of national economy and, as a result, are not obliged to provide the official statistics with any information concerning their activity.

1. Description and characteristics of the industry

1.1. Definition of the industry

According to the ISIC Rev. 4 the taxi operation is classified as a part of the class 4922 *Other passenger land transport*.

This class includes such other passenger road transport:
- scheduled long-distance bus services,
- charters, excursions and other occasional coach services,
- taxi operation,
- airport shuttles,
- operation of telfers (téléphériques), funiculars, ski and cable lifts if not part of urban or suburban transit systems;

This class also includes:
- other renting of private cars with driver,
- operation of school buses and buses for transport of employees,
- passenger transport by man- or animal-drawn vehicles;

However, this class excludes ambulance transport, see 8690.

According to the NACE Rev.2 the taxi operation is classified into the class 49.32 *Taxi operation*. This class also includes: other renting of private cars with driver

In the Central Product Classification (CPC) the taxi services (ISIC 4922/NACE 49.32) are classified under the class 6411 *Urban and suburban land transport services of passengers* with 9 subclasses[1]:
- 6411 *Urban and suburban land transport services of passengers*,
- 64114 *Local special-purpose scheduled road transport services of passengers*,
- **64115 *Taxi services*,**
- 64116 *Rental services of passenger cars with operator*,
- 64117 *Road transport services of passengers by man- or animal-drawn vehicle*,
- 64118 *Local bus and coach charter services*,
- 64119 *Other land transportation services of passengers, n.e.c.*;

The subclass 64115 *Taxi services* includes:
- passenger transportation services by motorized taxi within or between urban and suburban areas and
- non-scheduled airport shuttle services;

[1] 6 of them corresponds to ISIC 4922

These services are generally rendered on a distance-travelled basis and to a specific destination. Connected reservation services are also included.

This subclass does not include:
- scheduled airport shuttle services, cf. 64114
- chauffeur-driven car-hire services, cf. 6411
- man or animal-drawn taxi services, cf. 64117
- water taxi services, cf. 64129
- air taxi services, cf. 64242
- ambulance services, cf. 93194

In the Statistical Classification of Products by Activity (CPA 2015) the taxi operation services are classified as follows:

49	**Land transport services and transport services via pipelines**	
49.3	Other passenger land transport services	
49.32	**Taxi operation services**	
49.32.1	**Taxi operation services**	
49.32.11	Taxi services	This subcategory includes: - motorised taxi services, including urban, suburban and interurban - non-scheduled airport shuttle services These services are generally rendered on a distance-travelled basis and to a specific destination. This subcategory also includes: - connected reservation services This subcategory excludes: - man or animal-drawn taxi services, see 49.39.35 - water and air taxi services, see 50.30.19 and 51.10.12, respectively - ambulance services, see 86.90.14
49.32.12	Rental services of passenger cars with driver	This subcategory includes: - chauffeur-driven rental car services, wherever delivered, except taxi services These services are generally supplied on a time basis to a limited number of passengers and frequently involve transportation to more than one destination.

1.2. Specific characteristics of taxi operation market

Similarly to other countries in Poland the market for individual road passenger transport in Poland is not homogenous. The market consists of two segments: traditional taxi services and “ride-hailing” services.

The traditional taxi services are provided based on a telephone order placed through personal contact with a traffic dispatcher or by hailing a taxi on the street "on call." The fare is then charged according to the taximeter reading. However, the maximum level of prices (initial charge and price per km) are regulated by local governments. There are different types of companies which operate locally: taxi corporations serving a specific municipality or urban agglomeration, taxi driver associations and taxi drivers operating as sole proprietors (mainly in smaller towns).

The ride-hailing segment allows passengers to request rides on demand via a mobile app, connecting passengers with drivers using private vehicles. Platforms mediate between customers and drivers, allowing them to set a destination, pay in advance, and track their vehicle. The prices are established dynamically, depending on the location, demand for services, distance etc. Unlike a traditional taxi, the service can only

be booked through an app and drivers are often independent operators. Taxi services of this type have been available in Poland since 2012.

The "app-based" transportation segment comprises companies which offer intermediation services in major Polish cities, along with collaborating fleet partners (with fleet management) and individual drivers. The fleet partner offer appropriately equipped vehicles, technical support, and accounting and tax services. Moreover, there are also technology companies which provide modern digital solutions, such as computer and mobile phone applications tailored to the specific carrier's business model, automated call handling (dispatcher-less switchboard), and virtual cash registers. [2]

In Poland the taxi operation market is regulated by the law. In order to render taxi services special license is required which is issued by the local governments after fulfilling some conditions (i.e. driving license, no criminal records, lack of healthy and psychological contraindications to working as a driver). The turning point was when the ride-hailing services appeared on the market. For a long time this segment was not controlled in any way. However, in order to ensure the equal opportunities on the market, the ride-hailing segment was covered by law regulations in 2024.

1.3. Market conditions and constraints

The taxi transportation market in Poland has been developing dynamically. It is estimated that in 2023 total taxi market (both segments) in Poland reached USD 1.5-1.7 billion, of which 55-60% was for ride-hailing services[3]. Undoubtedly, the ride-hailing segment has experience dynamic growth and gained immense popularity, changing the way taxis are ordered and affecting the competitiveness of traditional taxi companies. Companies offering app-based rides often compete on price, resulting from lower operating costs compared to traditional business models, and a diverse range of services that adapt to diverse and evolving customer preferences.

Simultaneously, based on data available in the official statistics it was established that in 2023 there were 26 686 of enterprises with their declared core activity - NACE 49.32 *Taxi operation*. They employed about 31 thous. of persons (4.8% of number of persons employed in the division 49 *Land transport and transport via pipelines*) and generated almost USD 757 mln of turnover (1,1% of the net turnover in division 49).

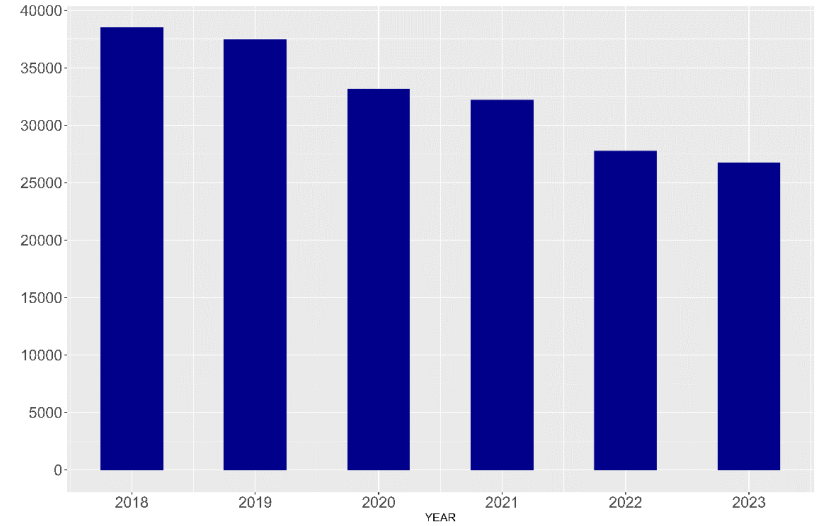
However, it has to be underlined that many companies, which render the taxi services, register their activity under other codes than 49.32 (by NACE Re.2). At the same time, in the population of enterprise which are registered under that code there are also fleet partners. Moreover, for some part of ride-hailing services segment data are not available as the Statistics Poland is not entitled to collect data from the companies (i.e. Uber or Bolt) which are not registered as entities of national economy.

All units operating on the taxi transportation market form a complex structure and, as a result, the precise separation of ride-hailing services from "traditional" ones is becoming increasingly difficult. Due to the diversity of their legal forms and types of business activity and high rotation of carriers it is impossible to determine the number of all entities operating in each segment of the taxi operation market.

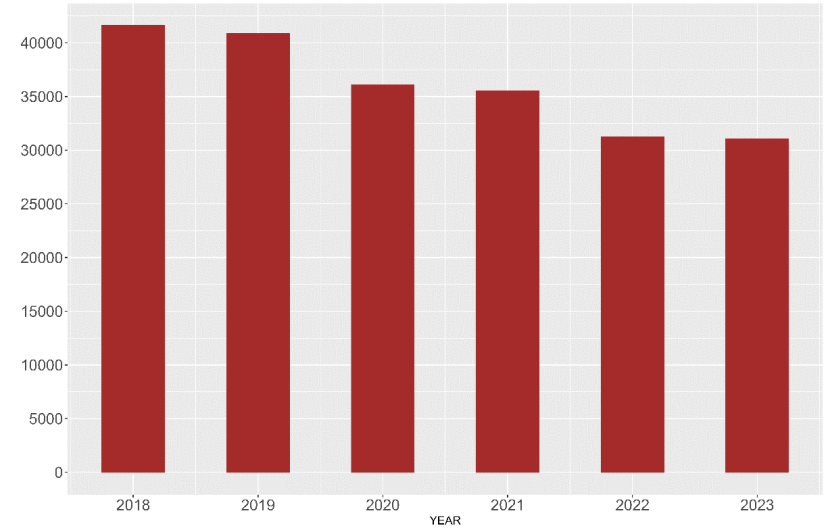
[2] Institute of Public Finance, Employers of Poland (2024): Road to nowhere - the costs of over-regulation of the cab transport market (https://pracodawcyrp.pl/storage/app/uploads/public/662/7a2/7f5/6627a27f54988761232315.pdf - only in Polish)

[3] TOR– Economic Advisory Team (2024). Smart regulations, efficient transport. Warsaw. p.1 (https://zdgtor.pl/wp-content/uploads/raporty/Bezpieczenstwo%20osobiste%20w%20przewozach%20taksowkami%20v.3%2011%20stron.pdf - only in Polish)

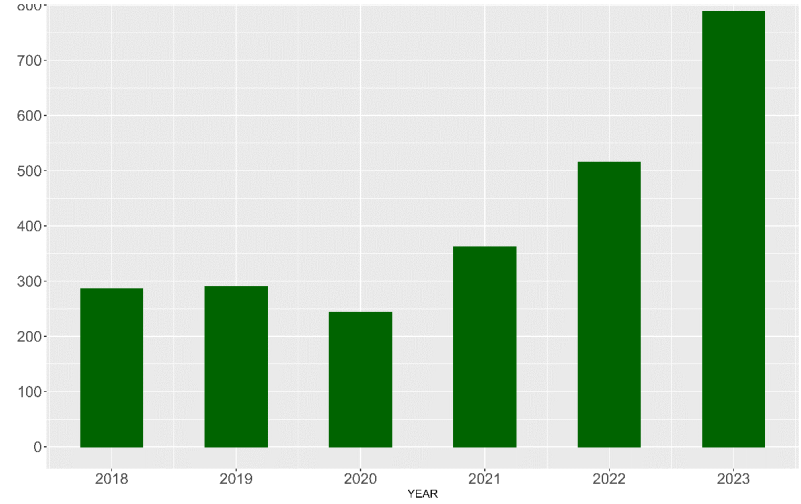
Graph 1 Number of enterprises with the core activity in NACE 49.32 in Poland in years 2018-2023*



Graph 2 Number of persons employed and self-employed in NACE 49.32 in Poland in years 2018-2023



Graph 3 Turnover in NACE 49.32 in Poland in years 2018-2023 in USD mln



Source: Statistics Poland, Structural Business Statistics
** Due to the change of statistical unit from legal entity to enterprise in 2021 data not fully comparable*

In period 2021-2023 the number of enterprises engaged in ISIC 4922/NACE 49.32 decreased by above 17%. In the same time the number of persons employed in these enterprises dropped by 12,6% while the value of generated turnover – increased by more than twice.

The population of enterprises classified to *Taxi operation* (ISIC 4922/NACE 49.32) is dominated by units with the number of persons employed 9 and less (including self-employed persons). In 2023 they constituted 99,8% of population classified into NACE 49.32. Simultaneously, those entities generated about 81% of turnover in NACE 49.32 and employed 93% of persons employed in that industry. However, the self-employed persons usually cooperate with the taxi corporations or taxi driver associations. Therefore, for the needs of the price statistics it is recommended to comprise these type of companies with the survey.

2. Measurement of SPPI

2.1. General framework

The Producer Price Index for Services (SPPI) is one of variables compiled in the European Statistical System (ESS) within the short-term statistics (producer prices – 130 201). In compliance with the requirements of EBS Regulation[4] data on SPPI are compiled quarterly for specified groupings by NACE Rev.2 and transmitted to Eurostat within 90 days after ending the reference quarter. The SPPI is obligatory for NACE 49 *Land transport and transport via pipelines* according to EBS regulation from 2021. Poland compiled index for NACE 49 since 1999.

Data on SPPI are used for deflating various nominal values in current prices, for example macroeconomic variables, turnover, revenues from the sale of products, etc. They are widely used in the national accounts statistics and business statistics. Moreover, data on SPPI are used when analyses of inflation are conducted.

2.2. Measurement issues

Indices of producer prices for passenger land transport services are compiled as a part of the regular survey of official statistics - *Survey on services producer prices*[5]. In the field of passenger transport the survey covers units with a number of employees of 10 persons or more, conducting both primary and secondary economic activities classified according to Polish Classification of Activities (PKD 2007 = 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;

For classes 49.31 and 49.39 services producer price indices are compiled on the basis of data obtained through the electronic application made available to respondents on the Reporting Portal of the Statistics Poland (C09 *Report on producer prices of transport, storage and telecommunications services).* In the case of class 49.39 data are also obtained from the price lists of enterprises providing public passenger car transport services. On the other hand, in the case of classes 49.10 and 49.32, the data come from price lists published on the websites of the Polish State Railways (PKP) and taxi corporations (TAXI). Data collection requires each time entering the website of a given carrier and manually entering the data into the IT system of the survey.

[4] Regulation (EU) 2019/2152 of the European Parliament and of the Council on European business statistics, repealing 10 legal acts in the field of business statistics (EBS-Regulation) together with Commission Implementing Regulation 2020/1197 laying down technical specifications and arrangements.

[5] according to *Statistical Surveys of Official Statistics Programme* - symbol 1.64.16;

In order to improve the process of data collection and processing in 2024 the Statistics Poland undertook the development works, which were carried out 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.*

As part of the work carried out, the first step was to analyze the structures of the carriers' websites in terms of both the information provided and the technical requirements enabling automatic data collection. Based on the conducted review the websites of 35 taxi companies were selected for further works, of which 29 carriers are currently covered by scraping.

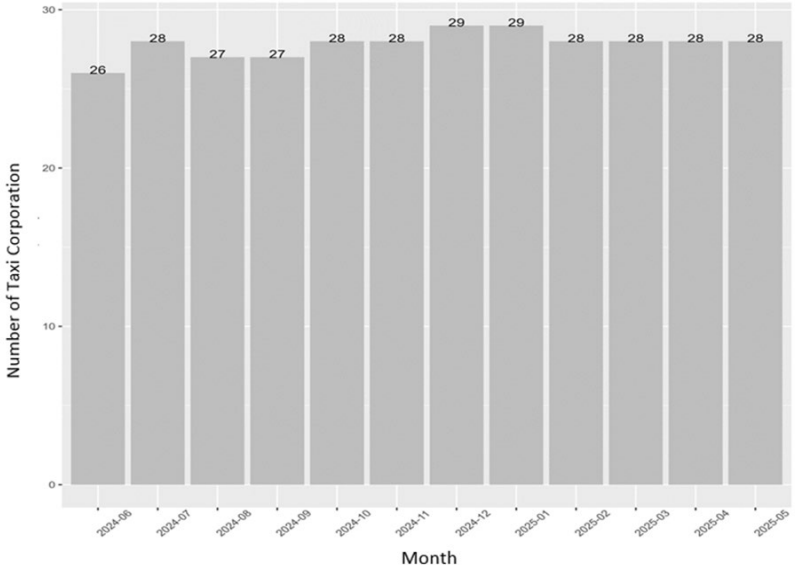
In order to ensure a uniform data set, it was agreed that for each surveyed corporation the following information would be collected on the daily basis: city, name of the company, email address of the price list, date of listing, initial charge and price per km in tariff 1.

Next a tool was developed to automatically collect data from carriers' websites. Data scraping began at the end of May 2024, and the results of the analyses presented in this paper cover the period from June 2024 to May 2025. The data scraping tool, known as the TAXI service, is installed on the departmental server and has built-in processes that automatically run at selected times of the day. In May 2025 the system obtained data for 28 corporations operating in 25 cities, while in 2024, 14 cities were covered by the observation.

Map 1 Number of taxi operators web-scrapped in individual cities in the period June 2024 - May 2025

Source: results of experimental works

Graph 4 Number of web-scrapped taxi operators in the period June 2024 - May 2025



Source: results of experimental works

In order to warehouse data collected via web-scraping, it was also necessary to create a database which consists of two elements. The first one is dedicated to data collection, while the second one, equipped with a user interface, enables the analysis of quality and completeness of the collected data. The data is exported once a month to Excel files and then processed in the R environment.

2.3. Methodology for calculating taxi service price indices

The methodology for calculating the taxi services price indices based on web-scrapped data was established, among other things, on the basis of the recommendations contained in the study *Modern technologies and new data sources in inflation measurement*[6] and includes the following 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.

After analyzing information on the functioning of the taxi services market, it was assumed that the price of a representative transport service includes the initial charge and the cost of a 10 km - ride under tariff 1. This is the average length of an average taxi ride in the city.

$$P_m = PO_m + 10\ km\ x\ Pkm_m$$

where:

P_m *- price for the service in a month;*

PO_m *- initial charge y month;*

Pkm_m *– price for 1 km transport at tariff 1 per month;*

In order to adjust the prices collected from the websites of carriers (prices payed by consumersthe gross price quoted is reduced by VAT. As data is collected on daily basis, for each company surveyed, the monthly price of the service is calculated as the geometric mean of daily quotations. Daily data collection is recommended due to the volatility of taxi companies in the database and incidental/temporary technical problems in accessing the websites of selected taxi companies. The adopted solution allows for a flexible

[6] J.Białek, M.Kłopotek, T. Panek: *Modern technologies and new data sources in inflation measurement* (https://bws.stat.gov.pl/BWS/gus_bws_70_nowoczesne_technologie_i_nowe_zrodla_danych_w_pomiarze_inflacji.pdf)

response to the lack of access to information on the prices of a given company and ensures the continuity of quotations.

In the next step, the methods for calculating the price index were considered, regarding the characteristics of dataset. First of all, the web-scraping gives us access to data on prices, but there is the lack of information about the turnover which is necessary to establish a weighting system. Moreover, the observed rotation of companies in the database makes it impossible to maintain a permanent panel of reporters, as is the case in traditional price surveys.

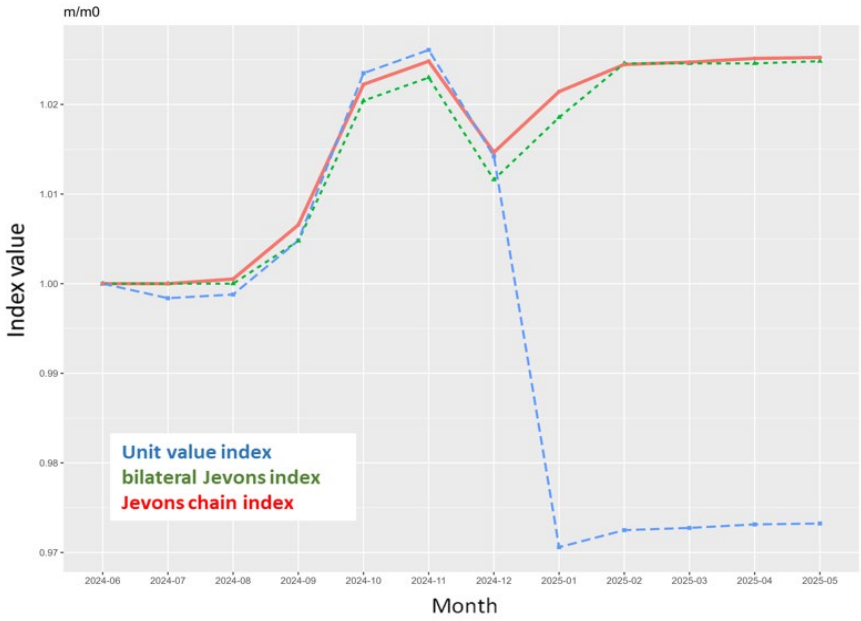
Three approaches to the index formula were analyzed:

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;

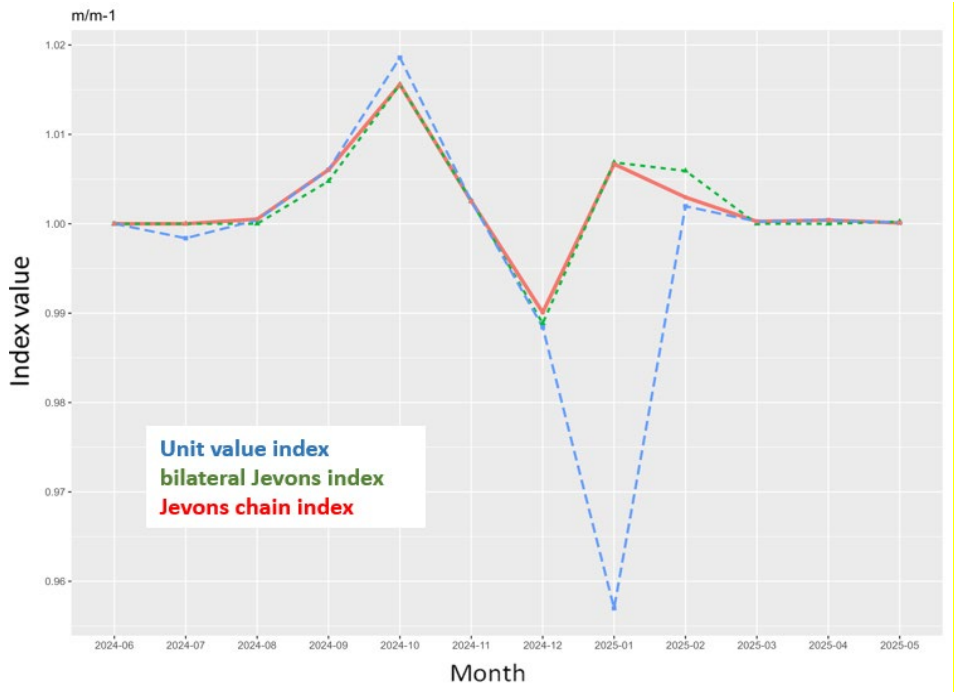
③ **the chain Jevons index** – for each period under review, price change indices are calculated in relation to the previous period for common observations, i.e., within a given city for the same corporations, and then the indices in relation to the base period are calculated using the chain method;

Graph 5 The experimental price indices for taxi operators activity by type of index formula – base month=100



Source: results of experimental works
** unweighted index*

Graph 6 The experimental price indices for taxi operators activity by type of index formula – the previous month=100



Source: results of experimental works

** unweighted index*

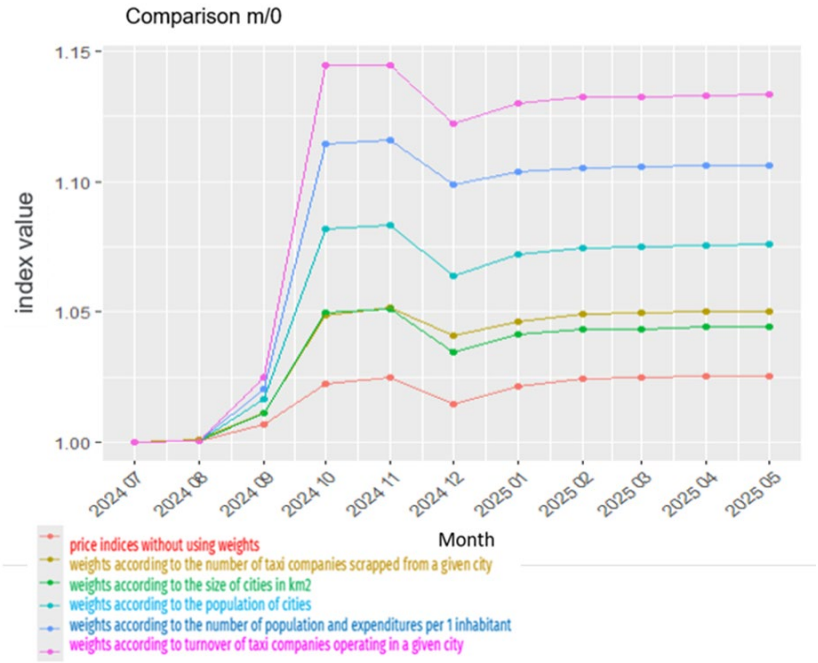
Based on the work carried out it was established that, depending on the approach used to construct the index, significant differences in price index values were obtained. However, the third approach was recommended as the solution, as it uses the largest number of price quotations and at the same time ensures the comparability of results between the periods under review.

The procedure for calculating price indices consists of two stages. In the first stage, the price for the city is determined for each month as the geometric mean of the prices quoted for corporations. Then, the price for the country is calculated as the simple or weighted geometric mean of prices for cities.

Therefore, an important element of the analysis was to determine the method of weighting price indices for cities. The following options for determining weights were specified:

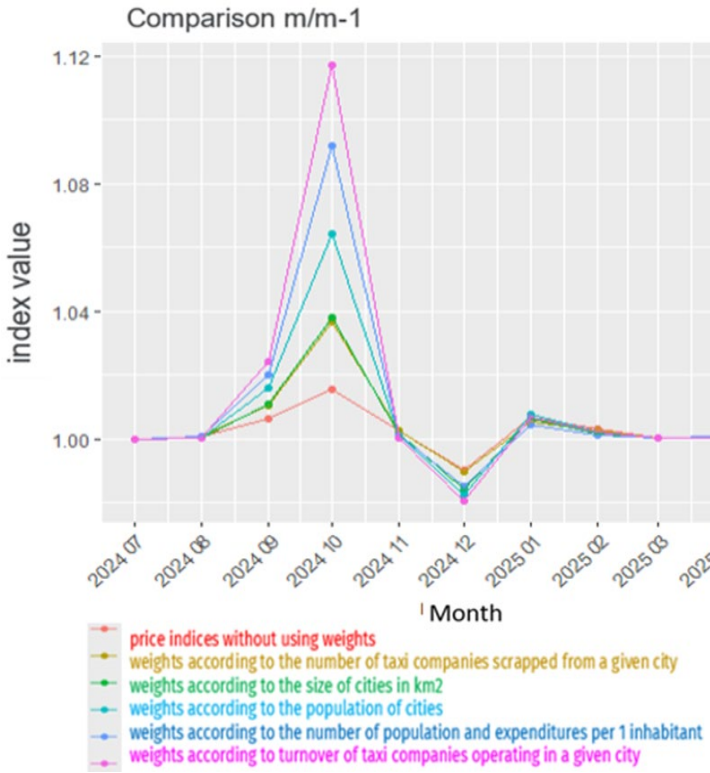
- according to the number of taxi companies registered in a given city,
- according to the population of cities,
- according to the size of cities measured in km 2,
- by population and per capita expenditure,
- by turnover of taxi companies operating in a given city,
- no weights – unweighted price index – geometric mean of prices for cities;

Graph 7 The experimental price indices for taxi operators activity by type of weighting scheme* – base month=100



Source: results of experimental works
* Jevons chain index

Graph 8 The experimental price indices for taxi operators activity by type of weighting scheme* – previous month=100

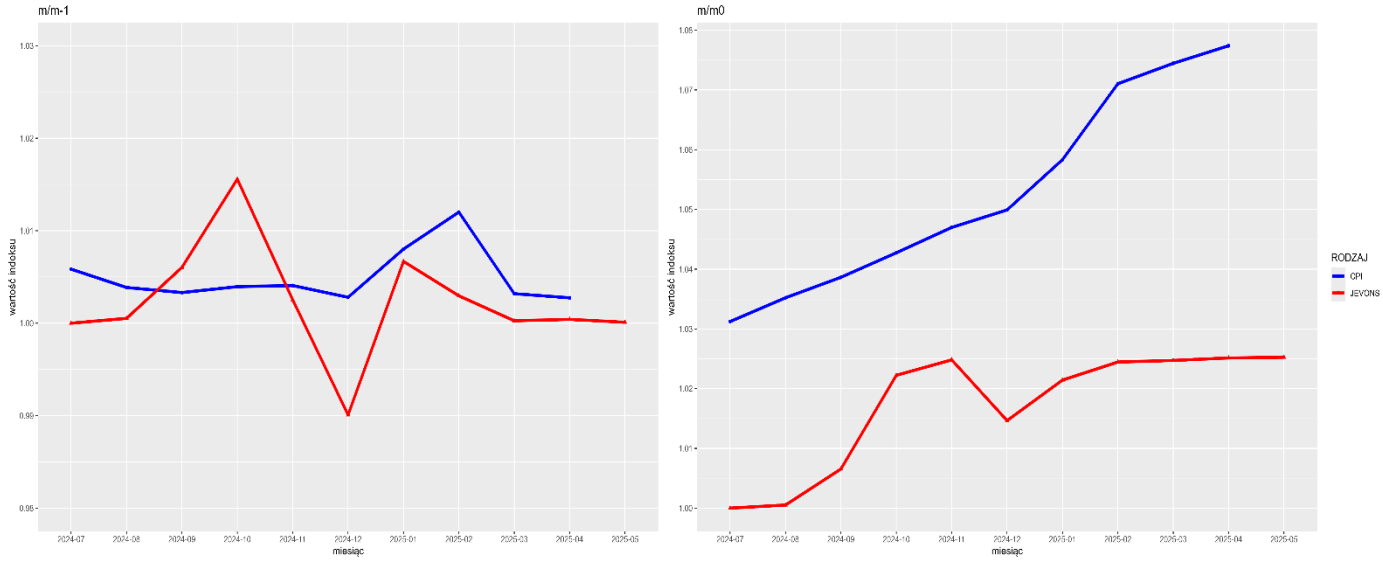


Source: results of experimental works
* Jevons chain index

A fairly important factor that has a significant impact on the values of the indicators is the method of weighting the results obtained for cities in the national indicator. Therefore, taking into account the short time series of aggregated data and the lack of taxi service representatives for all provincial cities in the surveyed population, it was assumed that at this stage, the most optimal solution would be a price index calculated using a simple geometric mean.

Due to the monthly frequency of scraped data, the results obtained were compared with monthly taxi service price indices calculated as part of the consumer goods and services price survey. However, the results from scraping, when compared with the traditional data source (even though they differ from each other), indicate the same direction of change. The differences in the values of the indices are primarily due to the greater number of price quotations in the regular survey than in web scraping at the current stage of work.

Graph 9 Comparison of experimental taxi service price indices with taxi service price indices (consumer price indices – CPI) for the previous month and the base month in the period June 2024 – May 2025*



Source: results of experimental works

3. Evaluation of measurement

The main conclusion drawn from the work carried out is a recommendation to use scraped data in price statistics. The benefits of this solution significantly outweigh the costs of its implementation in price statistics. However, it should be noted that the use of data collected from service providers' websites requires considerable effort during the implementation phase of new data sources in order to be able to effectively use the information resources collected in this way in subsequent stages.

Data from non-statistical sources, including in particular those collected through web scraping, are valuable sources of data for public statistics, offering access to a greater amount of information and, in the case of price statistics, also to a larger number of representatives. Data scraping is a more efficient and faster way of obtaining data than traditional statistical reporting. Processes programmed to automatically collect data from websites enable daily price quotations.

However, the implementation of new data sources poses numerous challenges for official statistics. First of all, it is necessary to develop a tool that will enable automatic data collection from service providers' websites and ongoing monitoring of this process. Furthermore, due to the need to update the tool and adapt it flexibly to market developments, it is necessary to provide a database administrator. The collected data must also be stored and processed in an appropriate manner, which requires sufficient server space.

Data obtained from web scraping also pose a methodological challenge for Statistics. Due to the specific nature of data obtained from websites, i.e. a large number of quotations, lack of information on turnover, or rotation of entities in the surveyed population, methodological solutions used for traditional data sources are not applicable here. It is therefore necessary to adapt the methodology for calculating price indices to these conditions. Efficient data processing also requires the implementation of IT tools that enable data processing and the calculation of price indices on the basis of scraped data, e.g. the R environment. This, in turn, necessitates the provision of widespread training on the above-mentioned tools for public statistics employees.