The 38th Voorburg Group Meeting on Services Statistics

26 October 2023

Cross-cutting topic

Croatian Experience in Updating Samples for SPPI, Including Initial Recruitment of Reporting Companies

Maja Dozet, Croatian Bureau of Statistics

Contents

Introduction

Services producer price indices (SPPIs) are essentially estimates based on a sample of prices of products produced by a sample of observation units. The purpose of this paper is to specify procedures in the Croatian Bureau of Statistics (CBS) for sampling and updating samples of reporting units for the calculation of the SPPI.

1. General aspects

Ensuring the representativity of services producer price indices has been a longstanding task for the Croatian Bureau of Statistics. Quarterly services producer price indices are important short-term indicators used to analyse business cycle trends in the service sector and as input for National Accounts statistics.

Development of SPPIs in Croatia started in 2008 under one of the European projects. CBS has already conducted a whole range of SPPIs as required by the short-term statistics (STS) Regulation and also several services which are not covered by this Regulation. Most of these activities were developed under different European projects.

CBS has gradually introduced surveys for different service activities and in 2022 eight pilot SPPI surveys were carried out. Service activities included in 2022 were: Land transport and transport via pipelines, except Road freight transport (as this survey was developed in 2013); Support services incidental to land transportation; Water transport; Air transport; and Other transportation support services; Accommodation; Food and beverage service activities; Programming and broadcasting activities; Real estate activities; Legal services; and Rental and leasing activities.

In 2023, surveys for three more service activities were developed: Technical testing and analysis; Other professional, scientific and technical activities; and Services to buildings and landscape activities, except Cleaning activities (developed in 2013).

For each service industry specific methodology has been developed considering its characteristics with regard to output, market with price mechanism and data availability, but respecting general methodological guidelines, the EU Regulation on short-term statistics and common conceptual framework of SPPI production developed at CBS.

Until now, there are 30 service activities which are collected by means of a web questionnaire, with prices for 4 200 items from 2 000 respondents in total.

The actual compilation of SPPIs in Croatia relies on B to All concept.

Data on service prices are mainly collected by sample surveys based on two-stage sample procedure: firstly, the observation units (enterprises) and secondly, the sample of services to be priced are selected on a questionnaire form.

2. Sampling process

Important steps in sampling are to establish and understand what the survey is trying to estimate, the limitations of the sampling frame, and the environment in which the survey will be conducted. Therefore, there is a need to collect information on the various markets concerned and also on main characteristics of the observation units involved in a specific service activity.

First step in this process is to meet with companies and professional associations as well as to consult specialised or not specialised literature (on the internet or in books and newspapers). This is done in CBS but not always at a completely satisfactory level, because of lack of time and also the reluctance of persons or organisations to provide information on the subjects. Due to the lack of other resources, in some cases the experiences of other countries are also used.

2.1. Selection of companies (responding units)

The best would be to select a sample of companies with "representative transactions" for each collection group within the set of companies, which provide services of the group. However, for the time being, the information is not available on what exact service a company has provided or even on what collection group services have been provided in a given company. Although the latter type of information may be obtained in future years by means of developing structural surveys on enterprises, which is not available for the time being.
The only information known on all enterprises of the universe is available in the Statistical Business Register (SBR). The information available on enterprises are main activities and the major secondary activity within the NACE Rev 2 classification. As to numbers, the only available information is the total turnover and the number of employees of each company.
The target population are business entities (legal units, tradesmen and free lances), which perform service activities that are proposed in the Requirement (sectors H to N of NACE Rev. 2).

2.2. Sampling frame

The representative sampling of reporting units requires comprehensive and up to-date sampling frame.

The sampling frame is determined by taking units from the Statistical Business Register according to the following criteria:

- Their main and secondary activity

- Turnover

- The number of employees

- Only active units are included

The sampling frame is also modified with collected information from other sources (companies, professional associations, literature, other surveys, …).

Statistical Business Register (SBR) is compiled and maintained using administrative sources (Administrative Business Register, Tax Administration, Central Register of Crafts, Commercial Court, Financial Agency, …) and statistical sources (Short-term Statistics, Structural Business Statistics and other surveys). SBR uses administrative concepts and transforms them to statistical ones.

Main variables that are used for stratification are number of employees and turnover. Administrative data on number of employees and turnover are selected primarily from administrative sources that can

provide data close to statistical definitions. If such data are not available, administrative data are used as proxy for statistical ones.

NACE code is another important variable that is provided from administrative sources. In majority of cases this information is assigned to units by administrative sources using methodology similar to the statistical one. In some cases, there is a need to replace administrative data with the ones assigned by statisticians.

SBR provides population for statistical surveys:

•SBS and other annual statistics–using main activity

•Short term statistics–using main and secondary activities

Table 1. SBR, 2020

Legal unit	Number of legal units
Legal units	340.950
Enterprises	245.028 (1.157 complex enterprises)
Local units	303.592
Enterprise groups	8.156
Foreign entities	

2.3. Sampling methods

The selection of observation units in CBS is done using either non-probabilistic methods or the probabilistic ones. The use of non-probabilistic methods introduces bias in the indices, which is very difficult to be estimated. The risk of bias is low in the case when the surveyed population is highly heterogeneous in the sense that a small number of large units contribute to the majority of the total of certain key characteristic (e.g. turnover). If, on the other hand, the surveyed population is more homogeneous, meaning that there is a large number of smaller units that significantly contribute to the total value of the population characteristics and, additionally, these smaller units might show different price behaviour, the bias introduced by the non-probabilistic approach can be quite significant.

In deciding how to select the sample of observation units, the degree of industry concentration is a relevant consideration. In CBS sampling methods used are:

a) Cut-off sampling – used when smaller number of large units contribute to the majority of the total of certain key characteristic (turnover)

b) Probabilistic sampling (for small enterprises) with census (for medium and large enterprises) – used when there is large number of smaller units that significantly contribute to the total value of the population characteristics.

A separate sample is developed for every SPPI service activity.

Table 2. Sampling methods for service activity surveys

NACE Rev. 2	Title of service activity survey	Sampling method	Sample size in 2023
H 49 ex 494	Land transport and transport via pipelines excluding Freight transport by road	Cut-off	65
H 494	Freight transport by road	Probability proportional to size with census	181
H 501	Sea and coastal passenger water transport services	Probability proportional to size with census	71
H 502	Sea and coastal freight water transport services	Cut-off	6
H 51	Air transport	One reporting unit (80% of total turnover)	1
H 521	Warehousing and storage services	Cut-off	12
H 522 ex 5224	Support services for transportation excluding Cargo handling services	Cut-off	54
H 5224	Cargo handling services	Cut-off	6
H 53	Postal and courier activities	Cut-off	20
I 55	Accommodation	Cut-off	77
J 58	Publishing activities	Cut-off	47
J 59	Motion picture, video and television programme production, sound recording and music publishing activities	Cut-off	42
J 60	Programming and broadcasting activities	Cut-off	25
J 61	Telecommunications	Administrative data source	
J 62	Computer programming, consultancy and related activities	Cut-off	124
J 63	Information service activities	Cut-off	18
L 68	Real estate activities	Cut-off	90
M 691 L	Legal services − lawyers	Cut-off	138
M 691 N	Legal services − notaries	Administrative data source	
M 692	Accounting, bookkeeping and auditing services; tax consulting services	Probability proportional to size with census	139
M 702	Management consultancy activities	Probability proportional to size with census	129
M 711	Architectural and engineering services and related technical consulting services	Probability proportional to size with census	160
M 712	Technical testing and analysis services	Cut-off	35
M 731	Advertising services	Probability proportional to size with census	79
M 732	Market research and public opinion polling services	Cut-off	28
M 74	Other professional, scientific and technical services	Cut-off	82
N 77	Rental and leasing activities	Cut-off	106
N 78	Employment activities	Cut-off	66
N 79	Travel agency, tour operator reservation services and related activities	Cut-off	86
N 80	Security and investigation activities	Cut-off	22
N 81 ex 812	Services to buildings and landscape excluding Cleaning activities	Cut-off	55
N 812	Cleaning activities	Cut-off	97
N 82	Office administrative, office support and other business support activities	Cut-off	158

2.3.1. Cut-off sampling

Cut-off sampling is in the CBS a strategy frequently used to select samples when the distribution of the selection variable (turnover) is concentrated in a small number of large companies. In this approach, a predetermined threshold is established with all units at or above the threshold included in the sample (selected with certainty) and units below the threshold level are not included (zero probability of selections). In CBS, cut-off samples are used targeting on 60% – 70% coverage on average in terms of turnover in every NACE class within the observed service industry, in combination with the number of employees (usually 20 and more). Sometimes, in some service industries, some NACE classes are extremely small in terms of output, classes that comprise less than 0.5 percent of total output in the observed service industry. Those classes are excluded from the sample. If this is not the case, then all classes within the industry are included for estimation.

Updating cut-off samples

The universe from which the samples are taken is changing over time. This poses a major problem, especially when establishments close while new ones open, new establishments are not introduced in the sample in real time, and the samples deteriorate. Considering that appropriate assigning of service activities is a rather complicated task, it is quite likely that the Business Register consists and certain misclassifications, over-coverage and under-coverage of enterprises.

As the data in samples are collected from the same companies on more than one occasion, general problems with such surveys are that the panel becomes depleted as the company stops producing and the panel becomes increasingly unrepresentative as time passes and the universe changes. Also, some companies may resent the burden of responding and leave the panel or provide poor quality data. All these problems cause bias.

Because of the mentioned reasons, CBS updates cut-off samples every year. This process is done by using an updated sampling frame from the Statistical Business Register (SBR) for all service activities. The most up-to-date list of enterprises with their turnovers and the number of employees is for year y-2.

The following activities are foreseen for the annual revision before carrying out the survey for the first quarter of the next year:

Updating the list of reporting units (RUs):

a. In principle, all units that are non-active, misclassified, in dormant or bankruptcy are excluded from the sample. Information on misclassified units is sent to the Statistical Business Register to be analysed and reclassified, if needed. RUs that did not cooperate have to be contacted inviting them again to cooperate.

b. Inclusion of the newly created RUs above the threshold. Data are obtained from the SBR.

c. Inclusion of additional substitute reporting units to maintain the same sample size.

2.3.2. Probabilistic sampling with census

Due to the significance of small enterprises in some service activities, sample design in those activities is set-up as a combination of probability sampling for small enterprises and census (take-all approach) for large enterprises. The class of large enterprises is selected by taking the units that have more than 20 employees or generate more than 7 million euro of turnover.

Random sample of small units is selected by using the usual design in the case of business surveys: the stratified one-stage sampling. The stratum of small enterprises is hence additionally stratified according to the NACE class (4-digit code) and the random sample is selected by using systematic probability proportional to size (PPS) sampling. Number of employed persons is used for implicit stratification. Turnover is used as a size variable in PPS approach.

Selection of units into SPPI surveys by using probability sample demands appropriate amendment of SPPI index compilation in terms of defining the weights. There are two types of weights in SPPI compilation: basic weights and modified weights.

Basic weights are weights based on turnover data from every reporting unit. Data on turnover are fulfilled in the questionnaire form in the first quarter of the year and are related to the turnover of the previous year. Final index for the activity could be calculated using only these basic weights (for cut-off sampling).

Modified weights are used when random sample approach is used for the selection of the observed units. These weights are obtained by combining basic weights and the random sample grossing-up factors calculated by strata. Grossing up factor for a stratum is calculated through the following steps:

- Sample design weights calculation. Sample design weights are determined as the inverse value of the inclusion probability. Since PPS sampling is used (with turnover as size variable), design weights are determined as the relative share of turnover for a certain unit inside the stratum, multiplied by the number of selected units in stratum.

- Non-response adjustment. Non-response adjustment coefficient is calculated by taking the ratio of sampled units to the sum of responded and out-of-scope units in each stratum.

- Calibration. With the calibration procedure weights are corrected in such a way that for the chosen auxiliary variables the weighted sums for the responded units equal to the known population values. In our case, turnover was chosen as the auxiliary variable. Because of the large number of out-of-scope units, the "frame turnover total" had to be adjusted to take into account this fact.

Compared to basic weights, modified weights increase relative importance of smaller, randomly selected units (inversely proportional to turnover) and of all units in strata with higher non-response and higher share of misclassified units.

Random sample of "small units" is introduced to reduce the bias, caused by the fact that one part of the population has zero probability of being included into sample in traditional sampling approach. On the other hand, with the introduction of probability sampling, each index has a certain standard error, but this error can (contrary to the bias) be (unbiasedly) estimated from the survey data itself.

Updating probabilistic samples with census

Creation of the new sampling frame and selection of the samples is based on two data sources:

1. Sampling frame that was obtained from the latest version of the Business Register by applying the same rules as the previous year. Following these rules, all the units that satisfy the following criteria are included in the frame:

   o Their main activity

   o They have at least 1 employee

   o Their turnover is not 0

   o Only active units are included

2. Information from the last year's survey. In addition to the above stated "old rules", the units from the last year's sample that turned out to be out-of-scope for our target population are excluded from the frame.

Selection of the sample is largely dependent on the sample from the previous year; the sample is in fact only renewed. The following rules are applied in this renewal procedure:

- Large units from the last year's sample that were designated as eligible units (also non-responding units) remained in the sample.
- New large units that appeared in the new frame are included in the sample.
- Small units from the last year's sample that responded in the survey were retained in the sample.
- Additional part of new small units was randomly selected in order to get the same sample size as we had last year.

Each year, the sample is only renewed for the units that did not respond or were out of scope last year. Therefore, with the allocation we calculate the number of new-small units that have to be added to the sample. The formula for calculation of this number (in each stratum) is given as:

$$n_h^n = n \cdot \frac{N_h}{N} - n_h^0; \text{ where}$$

N.. Number of all units in the population
N_h...Number of population units in the stratum
n.. Target sample size of small units
n_h^0...Old part of the sample ("survived" small units)
n_h^n...Number of new (small) units to be selected in each stratum

Every three years, a complete sample rotation of small units is done so the burden is shared among this kind of reporting units.

3. Recruiting establishments

Recruiting establishments means securing the cooperation of its staff, so that data will be of high quality. This task can be more effectively carried out via personal visits, especially for larger establishments, rather than via telephone calls or mailed notifications and questionnaires, but because of lack of time and work stuff, recruiting establishments is done mostly via e-mailed official notification.

A separate sample is developed for every SPPI service activity.

Launching of a regular survey is introduced by an invitation letter submitted to all units that are included in the survey. Along with informing the potential respondents on the purpose of the SPPI survey, data confidentiality and their role in its implementation, this letter is also used to ask the selected units for contact details of persons who will be in charge of completing the web questionnaire. After receiving the required information by e -mail, an automated message (directly from the application) is sent to those persons containing the user account data (username and password), which allows an access to the web questionnaire. After that, alert messages are sent every quarter to respondents with information on the beginning of data collection for the respective quarter and on deadlines.

4. Challenges

A major problem that occurred is difficulty to establish contacts with associations and large companies because of their refusal or lack of understanding of the topic. The climate in Croatia is such that companies perceive statistics as a big burden and are reluctant to deal with statistical questionnaires, especially small units. This is understandable as companies get a large number of questionnaires that must be filled in, not only for national statistics, and at the same time they must perform their primary job.
The Croatian Bureau of Statistics has not yet established the practice of prosecution for such reporting units (only in certain exceptional cases), although the Official Statistics Act provides for this possibility. At the meetings with some representatives of reporting units (those meetings that were very successful), it turned out that persons from professional associations or companies sometimes have problems with comprehending statistical concepts and the rationale of statistical production. This causes a certain communication problem: on the one hand, there is no sufficient knowledge about business and about the functioning of individual services; but, on the other hand, data providers, for instance, do not understand certain instructions or what is meant by a particular phrase in the classification.
The accuracy of the Business Register is a permanent problem when it comes to choosing a sample frame. The Business Register requires regular updates and quality monitoring. Efficient handling of different deficiencies of the Business Register (omission of units, misclassified units, falsely active units, delineation between main, secondary and ancillary activity etc.) and maintaining of a quality business register is not possible without a continuous cooperation between several departments in the Croatian Bureau of Statistics and participation of reporting units.
Despite all these shortcomings, the overall response rate for SPPI surveys is rather high – it amounts up to 80% on average.