

Use of Preliminary Data from the Tax Authority in EU-SILC

Mr. Rihard INGLIC

Senior Adviser, Statistical Office of the Republic of Slovenia

The main source of the data on income in the EU-SILC is the database kept by the Tax Authority. The Statistical Office receives these data twice a year; the first time in June N+1, when we receive preliminary data, and the second time in December N+1, when we receive final data.

The difference between preliminary and final data exists and we analyzed the two data sets: the number of receivers of a specific kind of income, the average amount of different kinds of income, etc. The analysis was done on raw data and on the data after all procedures used by the regular EU-SILC. At the end we calculated the poverty rate and some other main income indicators from both data sets.

1. Background

The EU-SILC is the reference survey for comparative statistics on income and living conditions [1]. In Slovenia the EU-SILC was conducted for the first time in 2005. It is conducted according to framework regulation and other regulations covering the EU-SILC.

The survey is output harmonized, which means that countries can collect the data in their own way and it is necessary only to provide all data from the regulation.

In the beginning Slovenia decided to collect as much as possible data with administrative sources.

The EU-SILC became the first sample survey where SURS used also administrative sources for collecting the data.

SURS has a right to get administrative data by legal basis and we have agreements with different institutions to this effect.

2. Timeliness

According to the regulation, Slovenia as a register country should deliver the data to Eurostat by 30 November N+2; for countries without registers the data should be delivered by 30 September N+1. In 2005, SURS had a delay with the delivery, because of different reasons, but after that we improved the date of delivery of final data to Eurostat from year to year. For the EU-SILC 2015 the data were delivered at the end of June 2016 (N+1). Because of earlier preparation and publication, it is still a challenge how to prepare the data earlier than today.

Current situation in Slovenia:

- Conducting the survey from January to June of year N
- Administrative sources – mostly in autumn of year N
- Tax Authority – final data in January of year N+1

We can see that the problem is the data from the Tax Authority, which we receive quite late.

The Tax Authority (FURS) collects the data according to the legislation on personal income tax. All dates are regulated. Employers must report about the incomes by 31 January for the previous year for all employees. After that FURS prepares decisions and they are sent to all persons in two parts. The first part is sent in March and the second in June. When a person gets a decision, the taxpayer has one month to complain to this decision. If the person does not complain, the decision becomes final. All final decisions are made in October or November and after that FURS has the final data. After that SURS receives final data for year N-1.

So far we have always used the final data by the EU-SILC and consequently we were able to begin with the data editing in January or February. At the same time we began also to conduct the survey for year N+1 and this means that we worked at the same time on two EU-SILC surveys, which is not in accordance with accelerating the process, as we have also personnel limitations.

In the EU-SILC 2015 we published the preliminary data in January and final data in July. The preliminary data did not include any data on income. In final data all income and poverty indicators were published. The income reference period was 2014, i.e. N+2.

Anyway, it is necessary to take into account that the EU-SILC in comparison to other surveys has a very long reference period for income – the whole year. Income is one of the most important topics in the EU-SILC and is not only a background variable.

3. Legislation

For using registers and administrative data SURS need a legal basis. The legal basis is the National Statistics Act, which in Article 4 stipulates that the reporting units shall be holders of official and other administrative data collections (records, registers, databases, etc.), and also natural and legal persons that are defined by the Programme of Statistical Surveys as data providers.

According to this Act, official collections shall be data collections, established by regulations or general acts of public power holders, on the basis of which certificates and public documents shall be issued.

According to the National Statistics Act, the Statistical Office has the right to get all administrative sources in Slovenia to use them for statistical purposes (Official Journal of the Republic of Slovenia No. 45/1995) [2].

FURS works according to the Slovenian tax legislation (Zdoh2), which defines the timeliness of all tax processes. SURS does not have any influence on this topic and we must accept it as it is. Payers of income must report about all incomes to FURS by 31 January for income from the previous year. FURS sends informative calculations of personal income tax to recipients of income in two batches – in March and May. Persons then have one month to object to the informative calculations. Complaints are usually because of change of tax relief, additional income for which FURS does not get the data (for example from abroad), mistakes, etc. After that FURS processes all these data and makes a final decision. FURS finishes with all these processes by the end of the year.

SURS receives preliminary data in June for the previous year. These data do not include all complaints and repayments/receipts for tax adjustment are also not included.

4. Data files from FURS

SURS receives from FURS several datasets. In the EU-SILC we use 3 datasets. In the datasets there are approximately 70 kinds of incomes. The datasets include approximately 1,700,000 persons of the just over 2,000,000 inhabitants in Slovenia. Children are mostly not included in FURS datasets, because they do not have any income. In the main FURS dataset for the EU-SILC there are several variables: kind of income, gross value, tax paid in Slovenia, social contributions, tax paid abroad and normalized costs. Net income is derived income, which is calculated during the data processing of all datasets from FURS. In data processing of whole datasets some extreme values are normalized and impossible values are edited.

Table 1: Ratio between some statistics by preliminary and final data – wage, 2012 and 2013

	Ratio	Difference in percentage points
Ratio 2012 sum gross final/sum gross preliminary	100.51	0.51
Ratio 2012 sum net final/sum net preliminary	100.46	0.46
Ratio 2012 receivers final/receivers preliminary	100.31	0.31
Ratio 2013 sum gross final/sum gross preliminary	101.01	1.01
Ratio 2013 sum net final/sum net preliminary	101.03	1.03
Ratio 2013 receivers final/receivers preliminary	100.63	0.63

Sources: SURS, FURS

We also checked some other income components, where differences were even smaller, but as regards contract work we found out some more critical cases:

Table 2: Ratio between some statistics by preliminary and final data for contract work

	Ratio	Difference in percentage points
Ratio 2012 sum gross final/sum gross preliminary	100.25	0.25
Ratio 2012 sum net final/sum net preliminary	80.19	-19.81
Ratio 2012 receivers final/receivers preliminary	98.09	-1.91
Ratio 2013 sum gross final/sum gross preliminary	101.00	1.00
Ratio 2013 sum net final/sum net preliminary	100.16	0.16
Ratio 2013 receivers final/receivers preliminary	100.08	0.08

Sources: SURS, FURS

As regards contract work we expected a larger difference than in any other regular income from work. The main difference is in the aggregate of income in 2012. In detailed analysis we found out that there are mistakes in preliminary data, because costs are not included in the file. This is a warning for SURS that mistakes can occur in preliminary data and that everything should be checked more carefully.

5. HY145 - repayments/receipts for tax adjustment

These data are not included into the preliminary data sets. Consequently, we cannot have at the end real income of household. We can ask ourselves if this component has a large impact on the data on disposable income or not. We checked if it would be possible to use the data from the previous year instead of the current year. We prepared an analysis to find the difference between data from year N and year N-1. This analysis could be done, because we compare the data for some years ago.

Table 3: Difference between 2012 and 2013 for repayments/receipts for tax adjustment

Difference	Frequency	Percent
Current year lower by EUR 5,000 – EUR 10,000	6	0.08
Current year lower by EUR 1,000 – EUR 5,000	264	3.65
Current year lower by EUR 100 – EUR 1,000	2498	34.5
Difference between years lower than EUR 100	503	6.95
Current year higher by EUR 100 – EUR 1,000	3667	50.64
Current year higher by EUR 1,000 – EUR 5,000	294	4.06
Current year higher by EUR 5,000 – EUR 10,000	9	0.12
Current year higher by more than EUR 10,000	0	0

Sources: SURS, FURS

6. Calculation of poverty indicators

Before calculating the indicators we made the effect size – so called Cohen's d test for average of equivalised income with assumption that the difference by mean of final and preliminary data is zero. We got the following results:

Table 4: Effect size preliminary data vs final data

	2012	2013
Number of units - final data	27 265	28 064
Number of units - preliminary data	27 265	28 064
Variance - final data	43 265 545	40 505 242
Variance - preliminary data	41 757 952	41 901 185
Standard deviation	6 520	6 419
Mean final data	12 639	13 485
Mean preliminary data	13 014	13 376
d	0.058	0.017
Regression	0.980297	0.998574

Sources: SURS, FURS

From the data we can see that the difference exists, but it is not large (d is less than 0.20 in both years). Anyway, it is necessary to use the data with caution, because we found out that the mean in 2012 is larger for preliminary data, but in 2013 for final data. Probably this is the effect of using data of HY145N from the previous year. We tested unweighted data.

The final aim of SILC is to calculate income poverty indicators. We would check the difference between the data from preliminary datasets and final datasets.

For 2012 we got the following results:

Table 5: Number of units by at risk of poverty status and kind of data, unweighted data 2012

Preliminary data	Final data		
	Poor	Not poor	Total
Poor	2,804	184	2,988
Not poor	106	24,970	25,076
Total	2,910	25,154	28,064

Sources: SURS, FURS

Table 6: Number of units by at risk of poverty status and kind of data, weighted data 2012

Preliminary data	Final data		
	Poor	Not poor	Total
Poor	261,282	13,137	274,419
Not poor	9,406	1,717,100	1,726,506
Total	270,688	1,730,237	2,000,925

Sources: SURS, FURS

The most important persons who cause the difference are those whose poverty status changed and they are on the diagonal from bellow left to above right. All together approximately 22,000 persons do not have the same status of poverty. The results for 2012 are:

At-risk-of-poverty threshold

final data: EUR 7,273

preliminary data: EUR 7,166

At-risk-of-poverty rate

final data: 13.5%

preliminary data: 13.7%

The difference is 0.2 of a percentage point

And for 2013 we got the following results:

Table 7: Number of units by at risk of poverty status and kind of data, unweighted data 2013

Preliminary data	Final data		
	Poor	Not poor	Total
Poor	2,854	361	3,215
Not poor	330	23,270	24,050
Total	3,184	24,081	27,265

Sources: SURS, FURS

Table 8: Number of units by at risk of poverty status and kind of data, weighted data 2013

Preliminary data	Final data		
	Poor	Not poor	Total
Poor	265,960	27,777	293,737
Not poor	24,976	1,691,438	1,716,414
Total	290,936	1,719,215	2,010,151

Sources: SURS, FURS

In 2013 for many more persons the poverty status changed if we used different types of data – preliminary vs final.

At-risk-of-poverty threshold

final data: EUR 7,111

preliminary data: EUR 7,379

At-risk-of-poverty rate

final data: 14.5%

preliminary data: 14.6%

The difference is 0.1 of a percentage point

At first sight we can say that the difference is not large, but we prepared a more detailed analysis according to the gender and age.

Table 9: At risk of poverty status by age and data status

		0-17	18-64	65+
2012	Below the at-risk-of-poverty threshold – preliminary	13.9	12.4	19.2
2012	Below the at-risk-of-poverty threshold – final	13.5	12.2	19.6
2013	Below the at-risk-of-poverty threshold – preliminary	13.2	13.1	22.7
2013	Below the at-risk-of-poverty threshold – final	14.7	13.0	20.5

Sources: SURS, FURS

In this case we can see that the difference is larger in some age classes and moreover the data are changed into different directions. In 2012 for example for youth the poverty rate is lower in final data than in preliminary data, but in 2013 it is higher in final data than in preliminary data.

6. Conclusions

If SURS used preliminary data, it is obvious that the process would be accelerated and data would be available earlier. We estimate that the data would be published approximately 3 months earlier. But we should test the quality of the preliminary data. It is necessary to make a compromise between timeliness and quality of the data. If we produced the data earlier, they would not cover all aspects of income indicators satisfactorily, although the indicators on total are almost the same. Problems occurred by different sub-categories.

It is not clear how many times in the case of using preliminary data they should be published. It is possible that we publish only the data with the preliminary data as final data from SILC, although we know that these data are not perfect. Another possibility is that we publish poverty indicators two times: the first time with the preliminary data and the second time with the final data. But in this case it is questionable how to explain the difference in the data to the public. The public usually does not understand statistical processes and in this case trust in SURS to produce reliable data could be lost.

If preliminary data were received only by policy-makers, it is questionable what they would do with them if they knew that final data would follow after a few months, which would not be the same.

References

[1] Eurostat (2015) EU-SILC 065 2015 operation (Version November 2014)

[2] Slovenian Parliament (1995, 2001) The National Statistics Act, English version available at http://www.stat.si/doc/drzstat/zakon_o_dsta_eng.pdf