**EUROPA 2020 LFS INDICATORS ON REGIONAL LEVEL IN POLAND.
HOW TO IMPROVE THE INDICATORS PRECISION WITHOUT THE INCREASE OF THE LFS SAMPLE?**

Hanna Strzelecka, Statistics Poland, H.Strzelecka@stat.gov.pl

Agnieszka Zgierska, Statistics Poland, A.Zgierska@stat.gov.pl

**Abstract**

*In case when there is a need to evaluate some phenomenon by selected characteristics or by territorial division, the often encountered problem is finding an appropriate data source which delivers reliable results.*

*Two of the five “Europe 2020” headline targets are monitored with EU Labour Force Survey (LFS) indicators. Moreover, several other indicators covering different EU policy domains are based on the EU-LFS results.*

*Every Member State has to disseminate comparable and reliable data at regional level for regional policies monitoring needs. How to do this when the data are based on a sample survey, like the LFS, and response rate has been decreasing over time?*

*The objective of the research work carried out in Polish Central Statistical Office (CSO) was division of main LFS indicators (not only “Europe 2020”) into three groups:*

*1) ready for publication at regional level;*

*2) maybe possible - the precision of which may be expected to be improved to the acceptable level after the introduction of the significant methodological changes into the survey;*

*3) not possible at all for dissemination due to a low precision.*

*For indicators from second group several simulation tests were done.*

*The two most important effects were achieved within the framework of the research work, which significantly have developed possibilities of the public statistics as regards the publication of the LFS regional data:*

*1) specification of the base of indicators (from the list being the object of the research work) together with the codes of precision/quality embracing the historical data;*

*2) elaboration of the variants of changes in the survey methodology, allowing to extend significantly possibilities for meeting the information requirements included in the scope of the research work.*

**Keywords**: labour market statistics, Labour Force Survey, sample, precision*.*

**1. Introduction**

Every Member State has to disseminate comparable and reliable data at regional level for regional policies monitoring needs. How to do this when the data are based on a sample survey, like the LFS, and response rate has been decreasing over time?

**2.** **The objective of the research work[[1]](#footnote-1)**

The objective of the research work was estimation of the possible historical data and working out changes in methodology of the Labour force Survey allowing obtaining representative estimates regarding the selected labour market and education indicators in order to determine areas for intervention and monitoring the effects of the carried out actions at the regional level, i.e.:

* selected indicators included in the “Europa 2020” strategy, at the voivodship level (NTS 2), i.e.:
	+ - percentage share of persons aged 18-24 years with education no higher than lower secondary (gimnazjum) who do not continue education or training in the total number of population in the same age group,
		- percentage share of persons aged 30-34 years with tertiary education;
* NEET indicator (young people Neither in Employment nor in Education and Training) and the number of this group of persons for the age categories 15-24,
15-29 and 25-29 years at the voivodship level of (NTS 2);
* basic aggregations of the data within the scope of the labour market at the level NTS 3, voivodship cities, or in case there was alleged no possibility for presentation of the results at the particular NTS 3 level, NTS 3 aggregations.

**3. Methodology of the carried out work**

On beginning of the research work the available information resources were described in a detailed way together with the methodology for determining outcome indicators from the LFS. The second step was calculation of the objective indicators and the analysis of their quality (precision) due to assess the possibility of further dissemination and utilisation.

The LFS results for the years 2010 – 2013 compiled as quarterly, semi-annual, annual, one-and-a-half yearly and biennial data were analysed. The basis of inference on the possibilities for the usage of the indicators obtained from the survey with the use of the hitherto methodology were estimates of the relative standard error (the coefficient of variation – CV) obtained as the result of calculation work.

There was jointly analysed precision estimates for:

* 8 indicators at the level NTS 2 (voivodship) for the quarterly, semi-annual, annual, one-and-a-half yearly and biennial results;
* 160 indicators at the level NTS 3 (subregion) for semi-annual, annual, one-and-a-half yearly and biennial results;
* 58 indicators at the level voivodship cities for semi-annual, annual, one-and-a-half yearly and biennial results;
* 1 indicator at the level of aggregated NTS 3 subregions for semi-annual, annual, one-and-a-half yearly and biennial results.

In order to facilitate synthetic inference the obtained precision estimates concerning particular indicators and breakdowns were symbolised by precision codes and then aggregated.

Therefore, performing the analysis required defining the following elements:

* defining principles for precision evaluation (coding) for single estimates;
* defining principles for aggregation of precision codes over time (for particular generalisation period included within the duration of the analysis, i.e. in the years 2010-2013 – e.g. estimates for eight semi-annual periods), in order to obtain a synthetic estimate at the indicator level, aggregation (breakdown) level, the duration of a generalisation period (e.g. semi-annual, biennial) and domain (e.g. a particular voivodship, subregion);
* defining principles for aggregation over domains in order to formulate synthetic conclusions concerning a given level of aggregation (breakdown) for a given indicator and the duration of a generalisation period.

The effect of the performed analysis was appointing for particular indicators the estimates concerning possibility of their publishing and utilisation. The indicators were divided into three groups:

* **ready** for dissemination at regional level;
* **maybe possible** - the precision of which may be expected to be improved to the acceptable level after the introduction of the significant methodological changes into the survey;
* **rejected**, which could not be included in any of the above mentioned groups, i.e. which were neither reliable nor expected to improve their precision to the acceptable level.

The analysis pointed many NTS3 level indicators important for the project objectives as included in the second group. To improve their quality and enlarge the scope of possible utilization of regional level data some methodological changes to the survey were proposed. The proposals constitutes some variants of the alternative methodology. They were tested by the simulation on empirical data in order to prepare final proposals of methodological solutions applicable and to be recommended for the Polish LFS.

The potential solutions tested covered:

* application of one-stage sampling design instead of the two stage one;
* modification of the way of stratification and allocation of the LFS sample;
* modification of the way of the results calibration.

Taking into account:

* the results of first stage analysis, which showed some NTS3 level indicators as difficult to estimate with good quality,
* great interest in obtaining separate indicators for voivodship cities and the rest of voivodship area at the NTS2 level claimed by potential users of the regional data,

it was decided to define additional level of aggregation between NTS2 and NTS3, which excludes voivodships cities (capitals of the voivodships) from the voivodship area.

Finally, the levels of territorial aggregation considered in the second stage analysis (simulation of the alternative methodological solutions) were:

* “maximum variant”: NTS3;
* “minimum variant”: NTS2+ – that means separate indicators for voivodship cities and their complements (voivodships excluding voivodship cities).

Two variants of alternative allocation of the sample ware tested, one focused on the NTS3 level estimates quality and one focused on the NTS2+ level.

**4. The effects of the research and applied methodological improvements**

Within the framework of this research work, a set of recommendations was elaborated concerning the changes in the survey method, defining the new methodology of sample selection and generalisation of the results recommended for implementation, which in case of its implementation should in future significantly increase the set of information from the survey concerning labour market and education, available at a lower than the hitherto one level of aggregation.

On the basis of the completed analyses and simulations, the team carrying out the project-related work, recommended the following solutions regarding sample selection and generalisation of the results :

* keeping the two-stage sampling scheme, with the appliance of the hitherto principles of creation of the first stage sampling units (JPS) and determining the number of dwellings selected from the JPS;
* keeping the surveyed sample size at the level similar to the current one;
* introduction of the method for stratification and allocation of the sample in strata defined as „NTS 2+”, assuming specification of the voivodship cities as separate strata from the hitherto voivodship strata and algorithm of the sample allocation in the strata aimed at obtaining the satisfactory generalizations at the level of a voivodship city and a voivodships excluding voivodship cities;
* application of the calibration of the results at the voivodship level, replacing the hitherto used calibration at the country’s level, with the appliance of the hitherto used demographic criteria.

The recommended variant of the survey methodology should allow obtaining the satisfying number of estimates at the level of voivodship city and voivodship with exclusion of voivodship cities for the majority of the indicators covered by the project range (including all indicators analysed at the second stage), without foreseeable increase in the survey costs and with relatively small organisational changes with limited burdensomeness. There was not achieved significant increase of the possibility for the data usage at the level NTS 3 (which would require a significant increase in the sample). Even the use of NTS 3 oriented allocation allowed obtaining only limited and incomplete set of indicators of satisfying quality at this level (without increasing the sample). Therefore, this variant of allocation was rejected and further efforts for extending of the LFS results at the NTS 3 level by direct estimation was not recommended. It is worth underlining that in this case of the recommended solution the expected effect was strong enough so it could not exclude the possibility that after their implementation the quality of estimates in case of indicators rejected at the first stage for the aggregation by voivodship cities, would actually reach acceptable level. Therefore, there was recommended carrying out the work targeted at ex-post assessment of the quality and precision of the data obtained after the survey implementation in accordance with the new methodology in case of the indicators rejected at the first stage for the aggregation by voivodship cities.

The recommended variant of methodological changes was selected on the basis of carried out analysis on the effectiveness of the considered solutions, taking into account organisational and financial criteria, which had a significant impact on the final choice. In case, when any financial and organisational possibilities emerged allowing increasing of the sample by 50%, it would be possible, on the basis of the solutions worked out within the framework of the project, to obtain estimation of the acceptable quality for a significant number of indicators at the level NTS 3, constituting the complete set of data. In such case, there should be considered the appliance of the method for the sample stratification and allocation described as “NTS 3”, which assuming continuation of the hitherto sample size was not feasible due to lack of possibility for a significant and bringing a real informative profit increase in the stock of accessible data for a subregion level.

These recommendations have started to be implemented in the Polish LFS sampling and weighting since Ist quarter of 2016. As a consequence of the methodological changes in LFS regional data time series have been broken, but they are now coherent with demographic regional data estimations.

**5. Further work and achievements on the regional estimates quality**

In 2017 Statistics Poland has started a new research which will be finished this year. It verifies the real effects of the adopted improvements on the quality of regional estimates and continues the previous work being focused on the same aims. In case of NTS 2+ estimates, the effects of the methodological changes are observed. In case of NTS 3 level, where the effects of direct estimation was found not satisfactory, some alternative solutions are sought.

It constitutes two main fields of work:

* continuing the previous research work by:
	+ verification how the newly introduced methodological solutions work in practice,
	+ assessment of their effect on the quality of indicators, especially on the NTS 2+ level, taking into account empirical data obtained with their use (not only simulation),
	+ adjustment and ‘tuning’ of the sampling methodology or even correction it (if needed)on the basis of empirical verification;
* elaboration and testing the methodology for indirect estimation of LFS indicators on the NTS 3 level with the use of small area estimation (SAE).

The first assessment of the effects on the quality of NTS 2+ was done using the LFS data for the whole year 2016 (four quarters). The changes in the sampling methodology had been implemented into the survey gradually because of the rotation sample scheme – by sampling of new subsamples beginning since Q1 2016. As the result the 2016 LFS data was obtained with the use of the temporary mixed methodology and the effect of the improvement was only partial. Because of this, the assessment of the effect using the 2016 data consisted of two parts:

* assessment of the quality of the real 2016 estimates,
* simulation of the final effects assuming full implementation of the new methodology (to be achieved after complete subsamples replacement).

The first results of the 2017 research seem very satisfactory. The quality assessment shows the significant increase of the quality of estimates at the NTS 2+ level for 2016 results, which should be strengthened after full implementation. The majority of objective indicators should be possible to obtain with acceptable quality (A or B codes) and many of them with the best quality (A code). Preliminary empirical verification of the performance of the new methodological solutions allow to conclude that there is no necessity to implement any major corrections in the LFS sampling and weighting. Some minor adjustment and ‘tuning’ may be appropriate e.g. because of changes in nonresponse rate and its pattern.

1. *Research work was carried out within the framework of the Technical Aid Operational Programme 2014-2020.* [↑](#footnote-ref-1)