

Proxy interviews in the Polish LFS – reasons and determinants, effects, possible solutions to deal with
Hanna Strzelecka, Tomasz Piasecki - Statistics Poland

Introduction

The research work concerning the analysis of the phenomenon of proxy interviews in the Polish LFS and the possibility of reducing its occurrence was carried out by Statistics Poland in 2017-2018 in the action entitled 'Investigating the effects of proxy interviews on key LFS variables and building approaches to reduce the rate of proxy interviews in data collection'.

The work carried out included the following issues:

- assessment the frequency and main areas of occurrence of the phenomenon (subpopulations affected),
- finding the determinants of proxy interviews,
- assessment of the impact on the values indicators based on the key LFS variables,
- assessment and analysis of factors, circumstances, activities that could potentially increase respondents' propensity to provide direct interviews,
- building, taking into account the results of this assessment, some solutions and proposals for changes in the survey, which could increase the frequency of direct interviews and testing these solutions in a pilot study.

In the analytical part, concerning the description of the phenomenon, its effects and determinants, the main methodological problem was to distinguish between the impact of proxy interview on the measurement of the variables surveyed (effect of proxy on indicators values, proxy bias) and the effect of how the respondents' characteristics (including variables surveyed) influence the frequency of proxy interviews (variables surveyed as proxy determinants). If the same variable can both be a proxy determinant and be affected by the measurement bias, the separation of the effects is not possible only on the basis of empirical data, without additional assumptions. The attempt at such a separation was made as a part of the work using statistical modelling under certain assumptions, but the results obtained are largely conditional on these assumptions.

Empirical analyzes of the occurrence of proxy interviews were conducted on the basis of the data from the Polish LFS for 2016. The calculations related to the use of more complex statistical procedures were made using SAS software.

The part of the work concerning the development of potential solutions which might increase the respondents' propensity to provide a direct interview and the percentage of such interviews in the survey (reduce the proxy rate) included two pilot surveys. The first one concerned the respondents' preferences regarding various forms of the interview and their opinions on solutions that could facilitate or encourage the direct interview. The second one was a test of potential solutions to reduce the occurrence of proxy interviews in the LFS survey.

The analysis of the determinants of proxy interviews

Over the last years (2012-2017) the share of proxy interviews in Poland has been decreasing, beginning from 40.9% in 2012 to 37.0% in 2017. During the same period the share of proxy interviews in the European Union showed a quite stable tendency, at the level of about 30%. In 2016, what is the reference year for more detailed analyses, the rate in Poland was equal to 37.2%.

The comparison of proxy rates in some subpopulations is the way to show the groups mostly affected by proxy interviews and the respondents' characteristics which determine them. These was determined as:

- Age (high proxy for young persons (up to 24 years) and the oldest persons (90 years and more),
- Sex (high proxy for men),
- Household size (the rate increases with the increase in household size),
- Status on the labour market (the highest proxy characterises the employed in the private sector, persons working full-time as well the unemployed),

- Disability (disabled persons with considerable level of disability characterised much higher share of proxy interviews),
- Participation in formal education and persons combining formal and non-formal,
- The type of work requiring longer working hours (thus causing difficulty as regards possibility to meet the respondent at home), involving to a much greater extent men than women.

The comparison of the rates in subpopulations does not always allow to determine whether a given factor or characteristic is a real determinant of the phenomenon, i.e. whether it directly affects the probability of a proxy interview. For example, proxy interviews in a very large degree concern the youngest age group. In this situation, the subpopulation in which young people are overrepresented may have a higher proxy rate, even if the characteristics determining it have no direct impact.

In order to obtain the assessment of the impact of particular characteristics purified from the effects of such coincidences and differences in the structure, logistic regression was applied. It explains the occurrence of a proxy interview using various characteristics of the respondent and the interview as explanatory variables. Obtained estimate of the parameter related to explanatory variables representing particular characteristics allow to assess the impact of this characteristic on proxy probability purified from the influence of other explanatory variables included in the model.

The respondent's characteristics included in the model are:

- sex
- age
- place of residence (voivodship and the type of place)
- education
- status on the labour market
- number of hours worked in the reference week in the main job
- participation in formal education
- participation in non-formal education.

Additional explanatory variables referring to the interview, included to purify the characteristics' effects from their influence are:

- survey wave,
- survey period (week).

The conclusions from this analysis, regarding potential determinants of the proxy interviews, are that:

- proxy interviews are observed with much higher probability in case of men than women,
- probability of proxy interview strongly declines with the respondent's age, with the exception of the oldest age group where it is higher than in case of slightly younger respondents,
- participation in formal education also strongly influences (differentiates) the observed proxy rates, but it is much less important determinant of the proxy interviews occurrence than age; the high frequency of proxy interviews for the people participating in formal education is mainly caused by coincidence with young age,
- probability of proxy interview is higher in case of the unemployed than economically inactive persons, and for both these categories higher than in case of the employed with very short working time and lower than in case of the employed working typical hours,
- probability of the proxy interview occurrence is negatively related to level of education and positively related to the number of hours worked a week, and the last relation is stronger in case of persons working shorter than 40 hours than in case of persons working more than 40 hours (for working more than 40 hours the effect is statistically insignificant).

In case of economic activity status, the number of working hours, and probably the level of education and participation in education, these conclusions may be to some extent biased because of the measurement error related to the proxy interview. The attempt to assess this bias was the subject of the next analysis.

Assessment of the effect of proxy interviews on the key LFS indicators

The majority of LFS indicators are the shares (percentages) of individuals with particular features (values of variables surveyed) in the population or some subpopulations. The observed value of the variable surveyed on individual level (person) is the function of the true value and the measurement error, which may be the error due to proxy interview. We can also show the observed value as the function of the individual personal characteristic and the dummy variable which indicates proxy interviews. When we use such a specification in logistic regression, explaining the variable surveyed, the parameter associated with the proxy dummy variable represents the proxy measurement error (proxy bias).

Of course statistical procedures does not distinguish between effects being the result of proxy measurement bias and caused from the influence on proxy probability. If we use this procedure with the same set of explanatory variables like in the analysis of proxy determinants, we could reproduce the same results. To distinguish it some assumptions are required. They are expressed in the set of explanatory variables. To truly obtain proxy measurement effects we should use explanatory variable which effectively explain the variable of interest, but we also believe they are not affected (or not significantly affected) by proxy measurement error. That should be characteristics we may believe being known to proxy respondents. This was for example the reason for splitting labour status into two parts: working (assumed as better known by proxy respondents, included into models as explanatory variable) and looking for a job in case of not working people (assumed as less known, not included).

As explanatory variables in the models were used:

- survey wave,
- survey period (week),

and characteristics of the respondents, considered as less affected by the proxy measurement error and treated within the applied model as not biased in this sense:

- sex,
- age,
- place of residence (voivodship and type of place),
- education,
- participation in formal education,
- the fact of working during the reference period (but not the full labour status),
- “typical” number of working hours in the main job strongly approximated (with the use of rough bands).

The conclusions from this part of analysis, regarding the proxy effect (measurement bias) were that:

- there were not observed the effects of proxy interview, which would significantly modify the obtained values of the analysed indicators¹, although for the majority of them the measurement effects were assessed as statistically significant; the strongest estimated effect regarding the indicator for the total population comprises 13% in relative figure and almost 0.3 percentage point in absolute figure;
- the concluded estimations indicate that measurement bias related to the proxy interview may
 - slightly (in absolute figure, in relative figure between a few and a dozen or so percent) underrate estimates of percentages of persons with significantly lower than the average and significantly higher than the average number of hours worked and the percentage of persons participating in non-formal education and life-long learning indicator;
 - slightly (for the unemployment rate in 2016 by about 0.1 percentage point, i.e. about 2%) overrate estimates of the number (percentage) of unemployed persons and unemployment rate;

¹ activity rate, employment rate, share of unemployed, unemployment rate, life-long learning, percentage of persons participating in non-formal education, early leavers from education and training, NEET indicator and percentage of employed with hours worked <10, <19 and ≥ 50.

- the strongest effects were observed in case of indicators related to non-formal education;
- as regards indicators concerning economic activity, the potential effect to the greatest extent may concern unemployment rate, although – as was mentioned earlier – it is not significant from the point of view of interpretation of obtained results; the effects of measurement regarding the number/percentage of the employed were not observed;
- obtained estimates should be treated as conditional in respect to assumptions made when developing the applied models; in particular, biased may be estimates of effects regarding the number of hours worked (due to a direct relation of the analysed indicator with the probability of obtaining a proxy interview) and underestimated may be effects concerning the employed (due to the model assumptions).

It can be concluded that the answers given by other household members (as proxy) do not worsen significantly the obtained results of the LFS, compared to potential direct answers, which is positive information. Such state of affairs is probably influenced by fact that in most cases of proxy – the answers are given by the closest persons from the respondent's family, hence the ones most familiar with his/her employment and educational situation.

Developing approaches facilitating the reduction in the rate of proxy interviews in collected data

The basic tool serving the purpose of development of such solutions and assessment of their effectiveness was preparation of the opinion questionnaire for the respondents regarding the method of the LFS interviews implementation (see the annex). Its objective was finding the respondents' preferences regarding (1) the survey implementation method with and without the interviewers and (2) the inclination to give a personal telephone interview by the respondents who were absent during the interviewer's visit (thus reducing the rate of proxy answers).

The additional motive for the action was the need for recognition whether the change in the way of respondents' selection (from the currently used sampling of dwellings' addresses into the sampling of particular persons); could cause on the one hand facilitating more efficient (valid) sample selection, by reduction of selecting 'empty' addresses, while on the other hand - reducing the percentage of the proxy interviews. The incentive here were the results of the analysis on proxy rates in the LFS survey in the EU countries, which indicated that the countries using for the survey implementation the sample of persons obtain decidedly lower rate of proxy interviews.

Collected answers to the opinion questionnaire delivered information regarding the preferred method of the respondents' participation in a statistical survey. The most often respondents indicated contact with the same interviewer who had visited their dwelling. Respondents were the least interested in phoning the interviewer in the statistical office for the telephone number specially left for this purpose. A relatively small percentage of persons declared willingness to self-interviewing through the online survey (in this case the most willing were young respondents).

Moreover, despite the fact that analysis on the answers to the questionnaire regarding opinions about personal participation in the survey indicated that the decided majority of the respondents (over 70%) was willing to give a telephone interview, these declarations were unfortunately not confirmed in the first test survey in which the interviewers obtained respondents' telephone numbers and phoned in order to carry out direct interviews – conducting direct interviews succeeded in case of relatively low percentage of persons (about 17%) from whom telephone numbers were obtained with their consent for a repeated contact.

Making telephone calls to persons absent from the household during the personal visit of the interviewer – increased the share of direct interviews, albeit to a smaller extent that it could be expected, whereas it increases the survey cost and burdening the interviewers' network; therefore, it seems not very real for permanent adoption into the survey organisation.

The tested approach consisting in a change in the Polish LFS organisation into carrying out the survey on a sample of persons instead of the hitherto used sample of dwellings, resulted in a significant growth in the direct interviews rate. During the first observation of the second test survey, the indicator comprised 89%, while in the 2016 LFS survey it stayed at the level of slightly around 63%. Therefore it is a significant improvement. Obtained in the test survey indicator oscillated around the value achieved by other EU

countries applying in the LFS organisation the sample of persons. Therefore, the test survey confirmed that such survey organisation facilitates obtaining much better results within this scope. Although it should be noted that the increase in the share of direct interviews obtained with the use of the sample of people and telephone interviews was at the expense of the decline in the response rate.

Moreover the appliance of the sample of persons in the Polish LFS would require working out the Sampling Frame for Social Surveys by the Statistics Poland in order to ensure compatibility of persons registered in dwelling at the level allowing selecting the sample of persons directly from the sample frame. It would be also desirable that the sample frame included also the information about the person's telephone number which would allow much larger use of the CATI method in the Polish LFS. There should be also considered other aspects, which were not examined within the scope of the project, including the cost of the changed survey organisation. Another important aspect that should not be forgotten is the problem with collecting the demanded variables concerning the entire household with which struggle countries applying the sample of persons in the LFS.

Therefore, taking into account the above conclusions, it is currently not recommended to change the way Polish LFS is organized. Respondents prefer to contact the interviewer they know, and relatively few respondents also volunteered to participate in the survey conducted via the Internet.

Admittedly, the change in the selection of respondents by introducing the selection of people instead of addresses of flats could contribute to an increase in the share of direct interviews in the data collected, however this would result in a decrease in the survey response rate, which is a key factor against such changes, the more so that the analysis showed that the influence of the proxy on the results of the LFS indicators is not so significant.

Annexe - opinion questions concerning the method of the LFS interviews completion

1	<p><u>In case of your absence at home during the interviewer’s visit, hence hindered contact, would you be willing to personally provide the answers to the questions of the labour force survey questionnaire by telephone?</u></p> <p>1. yes <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/> → 2</p> <p>2. no <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/> → 3</p>
2	<p>Which of the following methods of conducting a telephone interview would be the most convenient for you?</p> <p><i>Any number of answers may be chosen. In case of choosing more than one answer, please order them from the most preferable (3) to the least preferred (1). In case when a particular answer is not selected, please insert “0” for this position.</i></p> <p><i>Please insert adequate codes in the boxes next to the selected answers.</i></p> <p>1. you would be phoned by the interviewer who carried out the interview in your household <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>2. you would be phoned by another interviewer – the so-called teleinterviewer, the employee of statistical office <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>3. the interviewer who carried out the interview in your household would leave the number for a teleinterviewer in the Statistical Office, asking for you calling back in convenient term <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p>
3	<p>Please, state the reason for your refusal to carry out the telephone interview.</p> <p><i>Any number of answers may be chosen. In case of choosing more than one answer, please order them from the most significant (5) to the least important (1). In case when a particular answer is not selected, please insert “0” for this position.</i></p> <p><i>Please insert adequate codes in the boxes next to the selected answers.</i></p> <p>1. I do not trust telephone questionnaires <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>2. I am anxious about the safety of the data I provide <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>3. I do not answer unknown telephone numbers. <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>4. I prefer personal contact. <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>5. other (please specify?) <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p>
4	<p>Because we would like to make completion of the interview the least possible burdening for you, please do indicate the preferred method of the future survey implementation.</p> <p><i>Please order the answers from 4 to 1, where “4” means the most convenient interview method, while “1” – the least convenient. In case when a particular answer is not selected, please insert “0” for this position.</i></p> <p><i>Please insert adequate codes in the boxes next to each answer.</i></p> <p>1. interview carried out via the Internet – completing the questionnaire by yourself (on computer, telephone, tablet) <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>2. telephone interview <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>3. interview carried out directly with the interviewer during his/her visit in the household. <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p> <p>4. other form (please specify?) <input style="width: 30px; height: 20px; border: 1px solid black;" type="text"/></p>