Estimating and communicating uncertainty in official poverty indicators at regional level

Keywords: *Poverty Indicators, Measuring uncertainty, Sub-national estimations*

1 Introduction

Given the key role played by poverty indicators in designing and monitoring social progress in the EU, it is paramount to produce and communicate to the public measures of the associated inherent and unavoidable uncertainty of point estimates. Although several statistical organisations have started to invest in identifying ways to measure and communicate data uncertainty, this is only being done randomly, being more common the lack of the communication of data uncertainty. This may lead to a misinterpretation of the statistical results by the users. Within this framework, by referring to the EU-SILC Quality reports, the European Commission has adopted the new EU regulation on Social Statistics (2019/1700) which requires that countries should provide estimates of standard error along with the EU-SILC main target indicators (AROPE, AROP, several material deprivation and very low work intensity). However point and standard error estimates of these measures are reported in the national quality report by gender and age for the entire country. Yet, a crucial role in poverty analysis is played by sub-national indicators to be used for bench-marking and assessing the efficiency of regional policies [1]. Since high national poverty rates may be accompanied by concentration of poverty in specific regions or, on the contrary, by widespread poverty across regions, it is essential to estimate poverty indicators at regional level. Regarding how the measure of uncertainty is communicated, the numerical communication method used in the national quality reports could be confusing especially for layman users since it assumes a certain level of statistical literacy. However, till now specific guidance on how to communicate uncertainty has not yet been developed and no consensus has emerged as a general recommendation for communicating uncertainty among different target audiences.
The aim of this paper is to suggest methods for presenting uncertainty in the Eurostat national quality reports by considering different types of audience. In the national quality reports, poverty estimates are often shown in tables, with a numerical indication of their precision. However, point estimates and standard errors for income-inequality indicators such as Gini coefficient and S80/S20 are not reported. It is worth noting that clearly communicating uncertainty measures for these figures may not be trivial, due to the possibility that this information may be mis-interpreted by the general public. Therefore, the present study focuses on estimating and communicating standard errors for the at-risk-of-poverty rate (AROP) at regional level by using linearization method. As a case study, we consider the case of Italy by also providing practical suggestions for communicating uncertainty in the national quality reports.

2 Methodology

In order to estimate the variance of the AROP indicator we refer to the linearization approach. Among the alternative linearization methods, we based our variance estimates on the concept of influence function [2]. This method has been adopted

by Eurostat to produce variance estimates for the EU-SILC social indicators. The variance of the linear approximation can be used as an approximation of the variance of the non-linear indicator considered. Suppose θ is a complex non-linear indicator, the variance of an estimator $\widehat{\theta}$ of θ is estimated by:

$$\widehat{V}\left(\widehat{\theta}\right) = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum \left(z_{hi.} - \overline{z}_{h..}\right)^2 \tag{1}$$

Where $z_{hi.} = \sum_{j=1}^{m_{hi}} \omega_{hij} \cdot z_{hij}$, $\bar{z}_{h..} = n_h^{-1} \left(\sum_{i=1}^{n_h} z_{hi.}\right)$ and z_{hij} is the value of a linearised variable.
Even if the derivation of computational formula is more complex in some case, the linearization approach may perform better than the resampling methods [3]. It is worth nothing that we also applied bootstrap and jackknife methods but the results are not reported here due to the lack of space.

3 Results

The EU-SILC sampling design adopted by Italy is a two-stage stratified sampling, in which NUTS1 and NUTS2 are considered as planned domains in the sample design. Table 1 reports point and standard error estimates at national and NUTS1 level for Italy obtained by using the EU-SILC data for the year 2018.

Table 1: Estimated standard errors for AROP: Italy at national and NUTS1 level

AROP	Point	Rel. s.e.
National level	19.9	2.050
NUTS 1		
Northweast	12.8	5.31
South	32.8	3.28
Insular	35.7	5.55
Northeast	10.9	6.77
Central	17.2	5.02

As expected, uncertainty measures for AROP at NUTS1 level are higher than those obtained at national level. Since National Statistical Institutes (NSIs) have different users, this communication method, based on numerical indication, which is similar to the method used in national quality reports, may be confusing especially for layman users since it assumes a certain level of statistical literacy. It is worth nothing that clearly communicate uncertainty measures for these figures may not be trivial, due to the possibility that this information may be mis-interpreted by the general public. Indeed as in this case, information regarding standard errors or relative standard errors is shown without an explanation of the meaning of the uncertainty range [4].
The effects of uncertainty communication depend not only on the characteristics of the target audience and on the relationship between the audience and the communicator, but also on the topic or source of the uncertainty. Important differences

between individuals,including level of expertise, prior attitudes, numeracy skills, education level, might mean that the same communication of uncertainty affects people differently [5]. Therefore improving verbal communication should be a good practice in order to understand the "magnitude" of the uncertainty among the general public. By reviewing current practices adopted by NSIs in their annual national reports on poverty and living conditions, the approach adopted by the Office for National Statistics (ONS) could be considered as a best-practice. Regarding the UK, the ONS provides a detailed explanation of uncertainty measures such as standard errors, confidence interval, coefficient of variation and statistical significance and how they affect estimates from surveys used for producing official figures[1].
In addition, various uncertainty visual methods are used including error bars plots where each point estimates are visualised with a bar and the confidence interval is plotted as an interval on top of each bar. This kind of plot are particularly useful when standard errors are obtained using cross-sectional data and is also easily understandable for layman users. By considering the Italian NUTS2 regions, Figure 1 reports error bars for the AROP.

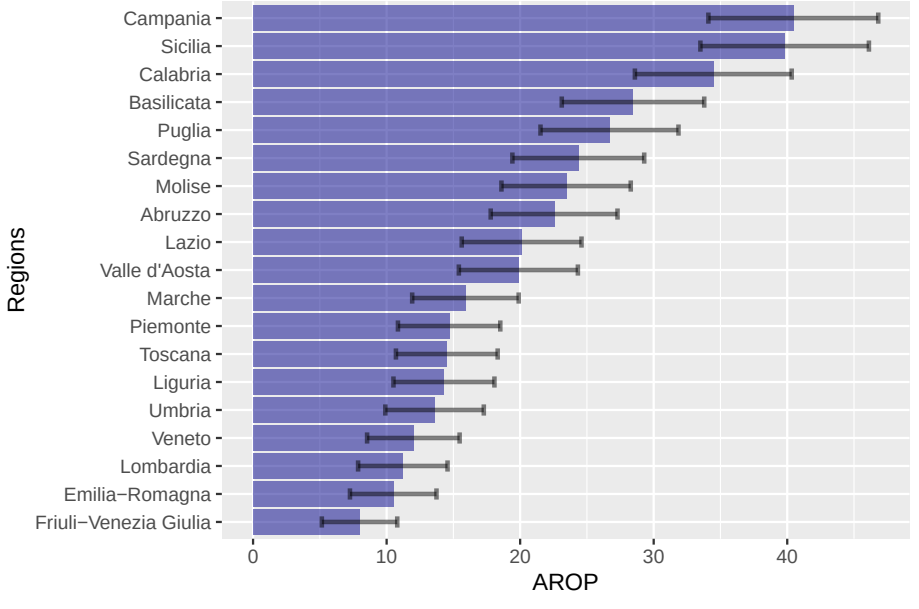

Figure 1: Estimated relative standard errors for AROP: Italy NUTS2

Since the available sample sizes at NUTS2 level become small, sampling error tends not only to be high, but also estimates of sampling error tend to be more complex and subject to high levels of variability. In order to overcome the problem of small sample sizes and produce regional estimates with reduced sampling error, various procedures can be implemented. It is advisable to improve size and unit allocation and/or using auxiliary information for computing small area estimation methods. Linearization seems to produce lower relative standard errors than re-sampling techniques for all the regions considered with the exception of Calabria and Campania where we found that the jackknife is better.
However, error bars do not provide indication of the underlying distribution of the number. For a further improvement of uncertainty communication a continuous-outcome visualizations approach could be used by referring to a probability distribution which describes a set of possible values for the estimated poverty indicators that are consistent to varying degrees with the data we saw and what our model assumes. For the sake of explaining uncertainty, a statistical model may be created that allows

[1]https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/uncertaintyandhowwemeasureit

pointing to intervals. This kind of uncertainty can be defined as the difference between the estimated and the true population value. The measurement of poverty is accurate, but not exact, since it is an estimate based on a sample of total population and it is therefore affected by sampling and non-sampling errors.

4 Conclusions

This paper focuses on the issue of measuring and communicating uncertainty for poverty indicators in Italy by using the 2018 EU-SILC survey. Information about the sampling variability of point estimates is essential when comparing poverty differences among geographical areas and when understanding if poverty rate has increased or decreased over time. A critical aspect in measuring uncertainties of poverty indicators is how to communicate them in a "comprehensive" way, in terms of capturing fully the uncertainties, but also in a "understandable" way so that different users and readers of these data correctly infer and interpret the uncertainties communicated to them. Increasing attention has been paid to this aspect in literature ([6], [5]). NSIs communicate uncertainty of poverty measures by providing a quantification of the magnitude of variability in their national quality reports. However, to the authors' knowledge, only few EU countries currently communicate measures of uncertainty when publishing new release on poverty and living conditions. To this respect, in this paper we provide standard error estimates at regional level and suggestions for communicating uncertainty in the national quality reports.

References

[1] Vijay Verma, Achille Lemmi, Gianni Betti, Francesca Gagliardi, and Mario Piacentini. How precise are poverty measures estimated at the regional level? *Regional Science and Urban Economics*, 66:175–184, 2017.

[2] Guillaume Osier. Variance estimation for complex indicators of poverty and inequality using linearization techniques. In *Survey Research Methods*, volume 3, pages 167–195, 2009.

[3] Vijay Verma and Gianni Betti. Taylor linearization sampling errors and design effects for poverty measures and other complex statistics. *Journal of Applied Statistics*, 38(8):1549–1576, 2011.

[4] Henning Griethe, Heidrun Schumann, et al. The visualization of uncertain data: Methods and problems. In *SimVis*, pages 143–156, 2006.

[5] Anne Marthe Van der Bles, Sander van der Linden, Alexandra LJ Freeman, James Mitchell, Ana B Galvao, Lisa Zaval, and David J Spiegelhalter. Communicating uncertainty about facts, numbers and science. *Royal Society open science*, 6(5): 181870, 2019.

[6] David Spiegelhalter, Mike Pearson, and Ian Short. Visualizing uncertainty about the future. *science*, 333(6048):1393–1400, 2011.

Acknowledgments: The EU-SILC microdata was accessed in the framework of the Comunikos project. The authors are grateful to all the people involved in the project for the fruitful discussions around this topic