EUROPEAN MASTER IN OFFICIAL STATISTICS
EMOS

Abstract
EMOS Master thesis competition 2023

'Estimating Spatial Distribution of the NEET Rate in Italy by Small Area Methods'

Author: Sara Sodini, University of Pisa

Keywords: NEETs; EU-SILC survey; Fay-Herriot model

1. Introduction

Studying the distribution of the NEET phenomenon [1] [2] has become more and more relevant over the last few years. The NEET label, which signifies that a young person is not in education, employment or training, first emerged in the United Kingdom towards the end of the last century. Due to the absence of a standardised definition at a global level and given the size of the category of reference, governments have adopted the NEET concept, applying it to the specific contexts that characterise their countries. Its use became widespread after 2010, when the European Union adopted the NEET rate as a reference indicator regarding the state of the younger population.

Formally, the NEET indicator is computed by Eurostat as a ratio where the numerator is the number of young people who are neither included in a work or study path nor in training, while the denominator is all young people in the population, having the same age as the reference group. It is important to distinguish it from the Unemployment rate, and we can find two main reasons: the NEET denominator is composed of all young people, while in the unemployment rate's denominator we find only the economically active population; the NEET indicator manages to capture not only the unemployed component, and therefore the active part, but also those who are inactive, even if this last category is at the same time the most heterogeneous, complicated and difficult to analyse.

Heterogeneity is a characteristic feature of the NEET population. The profiles found in the category, considering the 15-34 age group, are multiple and become clearer especially when the reasons that push young people to leave the training system and place themselves outside the labour market are analysed. According to the Eurofound [1], different macro-categories can be identified in order to describe the NEET category, such as the Unemployed, the Unavailable, the Discouraged workers and Other inactive.

With respect to the size of the phenomenon, Eurostat data showed that Italy had higher levels than the European average even before the economic crisis (18.8% in 2007 compared to 13.2% in EU-28).

The phenomenon increased most during the crisis (the rate rose to 26.2% in 2014 against 15.4% EU-28). According to the latest data available (2019, ISTAT), Italy has the highest percentage (22.2%) of young people in this condition in the entire euro area. This means that, even if in relative terms the percentage has slightly decreased each year, in absolute numbers the stock of young people is still very high. In fact, since 2013 the NEET quota has not fallen below 2 millions of young people. In addition, with regards to the last year, in Southern Italy, the incidence of NEETs was more than double (33.0%) that in the North (14.5%) and much higher than that recorded in the Centre (18.1%).

Extensive research has investigated the several and interlinked factors that cause the NEET status. The first set of factors are the ones related to each individual, such as the level of education, gender and all the other factors linked to the socio-economic status of the family. A peculiarity typical of Italy emerges with young people, as well as those already into their thirties, being dependent on their parents. Another Italian peculiarity, which certainly contributes to a higher NEET rate, is represented by the high share of the underground economy, where there is a large part of undeclared work. Last, there are also several factors at a macro level linked to problems associated with the economic and social reality of the country, such as the difficult path that young people are facing in the transition from school to work.

2. Objective

The main studies published on this subject, such us by Eurofound [1], as well as the most authoritative statistics produced on NEETs (e.g. those published by EUROSTAT or ISTAT), give us an in-depth, high-quality and, in many cases, internationally comparable information framework. However, these estimates come from sample surveys (Labour Force Surveys),that are representative at a macro-areas territorial level, such as countries and regions. Unfortunately, this does not allow us to go into a deeper detailed level (beyond those considered in the design of the surveys), and this weakens the cognitive value of the estimate, especially when we are moving from a theoretical plan to the planning and design of policies.

One of the objectives of this work is therefore to estimate the spatial distribution of the NEET rate at the Italian level, for those areas that are not planned in the stage of the survey design, such as the DEGURBA-Region level. In this categorisation the different regions are classified through the criterion of the degree of urbanization, thus identifying three main areas: cities, towns and suburbs and rural areas.

Since the typical direct estimation techniques do not produce reliable results for this level, the small area estimation methodology has been applied [3]. The data used to obtain estimates come from the European EU-SILC survey [6]. However, this is not recognised as the official source for obtaining

the NEET rate. Therefore, another aspect of this work has been the creation and subsequent estimation of the NEET rate in the Italian regions using EU-SILC data. An analysis was performed to verify whether starting from a different survey would make it possible to obtain a similar rate to the one that results from the official Labour Force Survey source, obviously applying the same definition that the National Statistical Institute (ISTAT) gives to the NEET category.

Another reason for the choice of the EU-SILC survey has been related to its possibility to estimate several important indicators, such as the at-risk-of-poverty or social exclusion (AROPE), employed to depict the poverty situation in Italy.

3. Methods

In order to measure the Italian territorial distribution of the NEETs, it was first necessary to reconstruct the variable itself within the EU-SILC data sets (2016). In particular, following a preparatory phase of the datasets, some variables were selected and through them it was possible to create the NEET dichotomous variable. To define the NEET category, the definition provided by ISTAT was used. In the 2016 data collection, 21,325 households and more than 40,000 individuals were interviewed.

One important adjustment that was performed in the whole dataset, was the re-calibration of the survey weights. In the EU-SILC file, every interviewed person has a weighting component, representing his or her probability of being selected in the sample and aspects such as the non-response factor. In order to obtain the real weight of a person living in a specific region, they need to be re-proportioned to the true population size in that region in the selected year.

In the last decades, the increasing demand for statistical indicators that described that the situation of local areas has led to the development of new methodologies. In particular, there was the need to obtain timely and detailed information also for those areas in which the sample size is not sufficient to obtain a reliable direct estimate [3]. An estimator of the parameter of interest for a given domain is said to be a “direct” estimator when it is based only on sample information from that specific domain. For example, the EU-SILC survey is designed to obtain reliable estimates at NUTS2 level in the European countries. Therefore, direct estimators can be computed only at regional level, and they cannot be used to compute the same parameter for an unplanned domain such as Provinces (NUTS3) or Regions by DEGURBA. Applying a direct estimation methodology, for those unplanned domains could lead to a not acceptable large standard error and very high coefficient of variations (CV) and as a consequence it is not possible to consider the result as reliable.

We have followed the guidelines of Statistic Canada on data reliability, with 3 main categories in interpreting coefficient of variations (CV) [7]. In this case, the aim of small area (domain) estimation (SAE) methods is to produce reliable estimators for the variable of interest under budget and time constraints. These methods try to overcome the problem of poor information for each domain by borrowing strength from the sample information belonging to other domains, trying to decrease the

coefficient of variations. The increase in efficiency of SAE is obtained using the information on units belonging to other areas considered geographically close or similar with respect to structural characteristics to the small area of interest. Small area estimation is a model-based method in which a model is usually constructed, where some dependent variables are expressed as a function of some independent variables. The models are classified into two broad types: Area Level which was used for the first time by Fay, R.E. and Herriot R.A.[4], and The Unit-level Estimator, used by Battese,G.E., Harter,R.M. and Fuller,W.A [5]. In this work the Fay-Herriot model has been applied since only area level summary data were available for the auxiliary or response variable.

In particular, the basic FH area level model rely on the following assumptions, and it is based in two subsequent stages. It is assumed that a basic unbiased direct estimator $\hat{\theta}_i$ is available, with θ_i being the true unknown value of the target parameter. Then, it is assumed a linear relationship between θ_i and a set of covariates X_i^T whose value are known for each domain of interest *i.* Combining the two model components, we obtain the linear mixed model i.e. the Fay- Herriot model:

$$\hat{\theta}_i = X_i^T \beta + u_i + e_i$$

Where u_i are the domain effects assumed to be normally distributed (as a white noise) and e_i the sampling errors, whit zero mean and constant known variance

Our final interest is to find a predictor at small area level with a lower variance that the direct one. In this case the Best Linear Unbiased Predictor (BLUP) of the target parameter can be obtained. In the work, in order to support the Fay-Herriot models, several covariates were selected from the ISTAT website “a misura di Comune”, such as Gross income per capita, Share of resident population by gender, etc.…

4. Results

As we can see in the Fig.1, in 2016 the italian NEET rate was 25.35%, that corresponds to more or less 3 millions of young people, that compared to the official one provided by ISTAT, 26%, is quite close but a little bit lower.

According to our estimates several southern regions were far above this percentage. The highest value was in Campania, followed by Sicilia, Puglia and Sardegna. More generally, all the regions in northeastern area were those with lower values, whereas the central regions were those more in line with the italian value. The situation depicted by the NEET rate, seems to trace the classical italian dichotomy between the southern and northern part of the country. We should not be surprised if the regions with the highest youth unemployment rates had also the highest NEET rates.

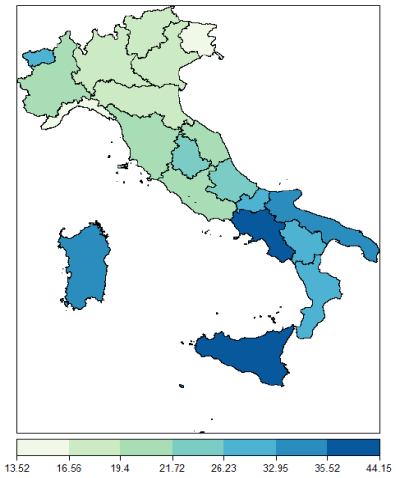


Figure 1 NEET Rate Direct Estimate

It was fundamental to check how reliable these estimates were. For this purpose, the values of the coefficient of variation were computed. In this case, the number of regions with a CV larger than the critical threshold of 16.6% was zero, that is all estimates seemed to be reliable. Confidence intervals were also computed. Since the larger is the sample size, the smaller and precise is the confidence interval, if there big differences in the sample sizes of the groups are found, these can affect the widths of the intervals and may give misleading results. On the opposite, if sample sizes are approximately the same, we can be confident that these differences are primarily due to different variation. In our case, the regions with a confidence interval larger than the others seems to be only three. Anyway, some of these intervals are overlapping to each other's. As a consequence, the difference in means could be statistically significant or not.

Another element of the analysis pertained the comparison between the Labour Force Survey (LFS) NEET estimate from ISTAT and the one obtained in this framework using EU-SILC data. One important difference is that EU-SILC considers a smaller age group, from 16 to 34 years, while instead the LFS has the possibility to include in the interview people from 15 years old. At the end, the NEET national EU-SILC estimate was quite close to the LFS, 25.35% and 26.02%.

For what concerns the different regions, in some cases the value were basically the same, such as in Piemonte, Veneto, Marche, Abruzzo, Sicilia, Sardegna and Puglia. In other cases, estimates differentiated as much as 1-3% from the actual real-world distribution such as in Lombardia, Emilia-Romagna and Lazio. Unfortunately, there were few cases in which the two were not so close such as for Valle d'Aosta (with EU-SILC is overestimating the LFS) and Calabria (with EU-SILC underestimating the LFS).

Focusing on a different territorial level, the degree of urbanisation within the regions, the results were more variegated. In particular, three labels were assigned to each region, "cities", "towns and suburbs" and "rural areas". The only exception was Valle d'Aosta, whose classification had only two levels (no urban areas). As a results, 59 sub-domains were obtained for the Italian peninsula, and

these domains were not planned in the survey. Therefore, we expected less reliability in terms of Coefficient of Variations (CVs). In fact, considering the 59 total domains, more than half of them presented a CV higher than 16.6%. Just the 29% of the domains presented a reliable direct estimator. For this reason, it was decided to apply a small area methodology, the Fay- Herriot model, in order to try to obtain NEET estimates with lower variability.

After applying Fay- Herriot model, the first important result was the gain in terms of variability. With the model 28 areas, representing the 48% of the analysed areas, had a CV below the 16.6%, threshold. The estimates are plotted in the figure below, with the FH estimates represented in the map on the right-hand side.

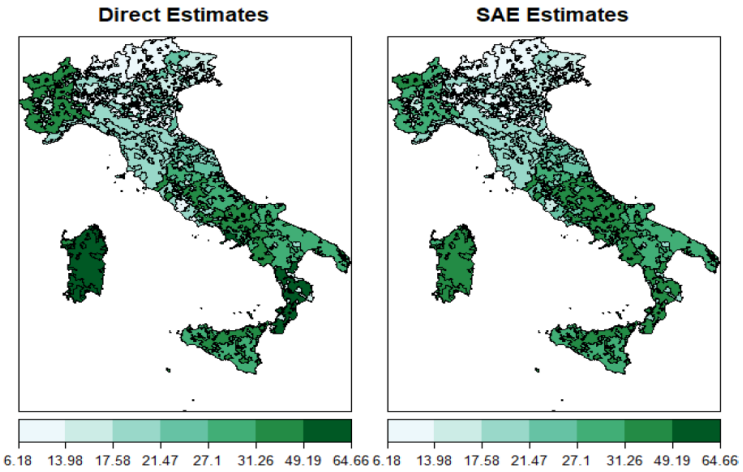


Figure 2 Direct and FH NEET total estimates

The gap between the North and South of the country is still present, but with this type of territorial classification, another aspect has emerged. Infact we can see that in sevaral cases the NEET rate was higher in those areas classified as rural, such as Sardinia or Calabria. Generally, it had been seen that less inhabited areas with fewer job opportunities

for young people and less educational services had also a larger share of NEETs. [1]

The same methods were applied also for a gender analysis. The NEET male rate estimated from the EU-SILC dataset was 21.64%, corresponding to about 1,338 million of young males. This average value was computed as the ratio between the NEET male population over the young male population of the same age group. The mean value for Italian young female was about 18.72%, corresponding to 5,951 millions of females all over the country. The female rate was therefore a little bit smaller with respect to the value of male population. A possible reason could be the slight gender imbalance for the young population sample, where the male observations are more frequent than the female ones.

A diagnostics analysis was also performed. Since the goodness of small area estimations relies on the quality of the specified model, it is appropriate to evaluate the implemented model, in order to

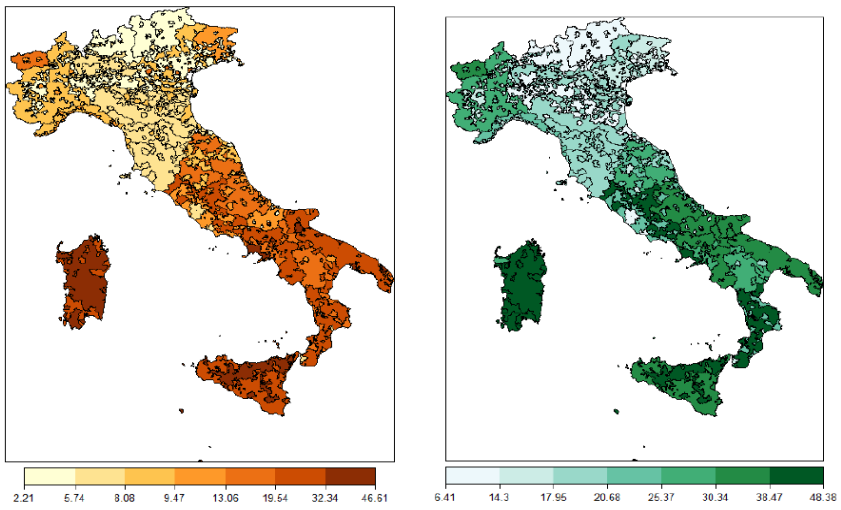
verify if some assumptions are violated or if there exists a potential bias. In addition, it was very important to check the correlation between the direct and Fay-Herriot estimates, that turns out to be highly correlated especially for the gender variable.

Finally, since the analysis is based on EU SILC data, it was possible to also compute the AROPE (share of total population at risk of poverty or social exclusion) indicator at both regional and regions by DEGURBA level. This is a very important indicator in the Europe 2020 strategy to

monitor the percentage of the at-risk population. It is a composite indicator, that is computed

as the sum of people who are either at risk of poverty or severely materially deprived or living in a household with a very low work intensity.

It was very important to address the NEET issue also from this point of view since the more detailed information it is possible to get for this category, the more it would be the possibility to apply targeted policy measures. Therefore, by combining the AROPE rate, one of its dimensions (economic), which is the at-risk-of-poverty rate, and the total NEET rate it was possible to outline the final scenario. Unfortunately, as we can see from Fig.3 the areas where there was a higher poverty rate and a higher AROPE were also those with a higher NEET rate, such as the southern regions of Campania and Sicilia.



(a) *At risk of poverty and social exclusion* (b) *NEET rate region by degurba*

Figure 3 AROPE and NEET, region by degurba

5. Contribution

In the thesis, in order to estimate the NEET spatial distribution throughout Italy, EU-SILC data were employed. Compared to the Labour Force Survey, which is the official source for estimating the NEET rate, the EU-SILC allowed to obtain reliable national data, which can also be used for country comparisons at European level. Using EU-SILC data it was possible to study the NEET phenomenon according to the specificity of the Italian territory, using the regions by degree of urbanization classification (DEGURBA), and to verify if some spatial patterns were identifiable. In addition, as the EU-SILC survey is one of the most important European sources for the estimation of poverty indicators (such as AROPE), this made it possible to better contextualize the NEET rate. Indeed, considering the NEET category also from this point of view allowed to better understand the potential criticalities to which they are exposed, and how it is important to delineate specific issues so that policymakers can tackle the problem providing targeted interventions.

References

[1] Eurofound. (2012). NEETs – Young people not in employment, education or training: Characteristics. Luxembourg: Publications Office of the European Union.

[2] Alfieri S., Rosina A., Sironi E., Marta E., & Marzana, D. (2015). Who are Italian “Neets”? Trust in institutions, political engagement, willingness to be activated and attitudes toward the future in a group at risk for social exclusion. Rivista Internazionale di Scienze Sociali, n. 3, pp. 285-306.

[3] Molina, I., & Rao, J.N.K. (2015, (Second Editon)). Small Area Estimation. Wiley Online library.

[4] Fay, R., & Herriot, R. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, 74(366), 269-277. doi:10.2307/2286322.

[5] Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association, 83(401), pp. 28–36.

[6] Eurostat. (May 2017). Methodological Guidelines and Description of EU-SILC Target Variables.

[7] Statistics Canada. (2016). Guide to the Labour Force Survey. Technical report, Statistics Canada, Catalogue 71-543-G, available at http://www.statcan.gc.ca/pub/71-543-g/71-543-g2016001-eng.pdf.