Mapping poverty at the local level in Europe

A consistent spatial disaggregation of the AROPE indicator for France, Spain, Portugal and the United Kingdom (*)

Alberto Díaz-Dapena[1], Esteban Fernández-Vázquez[2], Fernando Rubiera-Morollón[2], Ana Viñuela[2]

[1]University of Leon, Spain
[2]Regional Economic Analysis Laboratory– REGIOlab, University of Oviedo, Spain

Abstract

In the EU, territorial inequalities in terms of income and poverty have been broadly analysed at the national and regional levels. However, mainly due to the lack of reliable data, very little attention has been paid to territorial inequalities within European regions, i.e., at a more local level, such as in metropolitan areas, cities or neighbourhoods. This paper proposes a methodology to disaggregate official regional poverty figures into poverty indicators for smaller spatial units, mainly local administrative units. For each country, poverty figures at the regional level from household surveys are combined with microcensus data that contain details on the local entities of residence to disaggregate the regional poverty indicator. In contrast to previous methodologies, our proposed technique guarantees consistency between the local poverty estimates and the regional poverty figures through a second step that adjusts the initial estimates based on generalized cross entropy. The procedure is applied for four European countries: France, Spain, the United Kingdom and Portugal. The resulting local estimates provide an intraregional map of poverty and some insights into the particular behaviour of the capital regions and the disparities between city centres and their surrounding areas.

Keywords: territorial inequalities, poverty at small spatial scale, spatial disaggregation, entropy econometrics, European Union.

(*) This work is based on the results obtained in the European IMAJINE project (http://imajine-project.eu/), particularly on the report on inequality indices at local level (D2.4). The IMAJINE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 726950.

1. Introduction

In spatial economic analysis, great attention has been paid to the improvement of the estimation and inclusion of spatial dependence and spatial heterogeneity. This fact contrasts with the limited attention paid to the relevance of the spatial level of analysis. It is true that under the neoclassical framework, it is assumed that decreasing returns are the theoretical fulfilment of all spatial constructs and all levels of spatial disaggregation. However, theories of urban and regional economics and the new economic geography framework (Krugman, 1991; Krugman and Venables, 1995 and Fujita and Krugman, 1995) underline the local dimension of agglomeration economies and other sources of spatial disparities (see Fujita et al., 2001). In addition, a parallel body of literature on the modifiable areal unit problem (MAUP) explores the consequences of aggregating local information within larger regions or changing the groupings of regions (see Gehlke and Biehl, 1934 and, more recently, Openshaw and Taylor, 1979, and Openshaw, 1984). According to this literature, aggregated data could conceal heterogeneous intraregional patterns.

Since the international Great Recession, there has been more emphasis on providing a more comprehensive picture of the social dimensions of regional inequalities by means of indicators on, among others, well-being, happiness, poverty and social inclusion (see, for example, the studies of Milanovic, 2002 and 2005; Piketty, 2014; or Davies et al., 2017). That said, justified by the availability of official EU data, the analysis of regional disparities in terms of income, unemployment or any other socioeconomic indicator has been carried out at the level of NUTS 2 regions (see Fratesi and Rodriguez-Pose, 2017). Nonetheless, as Ballas et al. (2017) highlight, "the real social divides within Europe are more often within states rather than between them", that is, between regions belonging to the same country. Even before the 2008 economic crisis, regional inequalities within EU countries had increased by approximately 15%, while inequalities between countries had fallen by 45%, according to Heidenreich and Wunder (2008), with similar results obtained by Duro (2004) and Puga (2002). However, existing studies show how, after the Great Recession, income disparities between EU countries either increased (Dauderstädt and Keltek 2014) or remained rather stable (Darvas 2016), with inter-regional (or within-country) income inequalities accounting for 85% of the widening EU-wide gap by 2015 (Vacas-Soriano and Fernandez-Macías 2018). Regional inequality in Europe has increased in the last decade not only between regions but also within regions. As Ramos and Royuela (2014) point out, within-regional inequality has increased in 29 out of the 39 regions analysed, with "differences in intra-regional inequality higher than between countries".

Intraregional socioeconomic inequality has received some attention in countries where it may have become a driver of support for political attitudes that generate conflict and divide societies, exemplified by the result of the Brexit referendum in June 2016 (see Rodriguez-Pose, 2017) and the rise of populist parties across more than 63,000 electoral districts in the EU-28 (see Dijkstra et al., 2019). Combining individual and territorial factors, the above studies underline the existence of large social divides *within* regions that partially explain the most recent electoral results and justify the need to extend the scope of analyses to the *local* level (the LAU 2 level, in EU nomenclature).

Due to a lack of reliable local data and difficulties in accessing data at a small spatial scale, the literature on poverty and inequality (and growth) has mainly focused on creating indicators that can capture the complexity of the poverty phenomenon instead of trying to examine the scope of the phenomenon beneath the regional level (see Węziak-Białowolska 2015). However, European regions are geographically wide and economically heterogeneous, and their average regional poverty figures – regardless of the indicator chosen – mask large intraregional disparities. Understanding local differences in poverty and social exclusion is essential to adequately identifying spatial poverty concentration, designing local policies à la carte, and discerning the determinants of poverty and mitigating its consequences[1].

Using the AROPE ('At Risk of Poverty and social Exclusion') index as a measurement of poverty, as it is the one used by Eurostat in the EU-SILC (European Union Survey on Income and Live Conditions) dataset, this paper has the objective of dis-aggregating this multidimensional poverty index, available only at the regional level, and generating *local* AROPE indicators to show the existence of large poverty disparities within EU regions. Differences in poverty incidence between urban and rural areas or central and peripheral areas within the same region will thus be revealed.

The methodology proposed in this paper is inspired by previous literature such as Elbers et al. (2003) and Tarozzi and Deaton (2009). The point of departure of these studies was the combination of household surveys containing detailed economic information that is only reliable at an aggregate spatial scale with census data that are reliable at a small scale but do not contain economic indicators. Applications of this idea have been carried out by the World Bank to map poverty and inequality at the local level in countries such as Cambodia, Mexico, Morocco, South Africa and Uganda. Additionally, a similar methodology was applied in some of the 20 European countries covered in Copus et al. (2015), in which country microdata were available but AROPE figures were estimated for NUTS 3 regions only. This methodology, however, does not guarantee consistency between the disaggregated (sub-regional) estimates and the aggregate (i.e., regional or national) figures contained in the household surveys. The novelty of our proposal is that it applies a second step to the initial estimates to adjust them and make them consistent with regional and national aggregates. In particular, we propose using the Bernadini-Papalia and Fernández-Vázquez (2018) estimator based on generalized cross entropy (GCE) to adjust to observable aggregate estimations obtained through the Tarozzi and Deaton (2009) methodology. France, Spain, the United Kingdom and Portugal, where similar census information can be found, are used to illustrate this approach.

If local poverty estimations confirm the existence of large intraregional disparities within EU regions, as would be expected given the existing income disparities, it would prove the need to perform differentiated policy analysis and treatment depending on the scope of the governmental activities concerned and to raise the alarm about the existence of local areas of poverty within regions.

[1] Efforts to measure poverty at the local level in Australia were made by Miranti, McNamara, Tanton and Harding (2011) using alternative methodologies.

The paper is organized as follows. In Section 2, a description of the alternative approaches to estimating local data from other sources is presented. After we describe the standard *interpolation procedures* and the more advanced *microdata-based techniques*, the novel procedure proposed in this paper is explained. Section 3 shows how our methodology can be applied to estimate local AROPE figures for four countries in Europe: France, Portugal, Spain and the United Kingdom. Then, Section 4 is devoted to briefly discussing the resulting local estimations and generating detailed local poverty maps for those countries. The main conclusions and policy implications are presented in the final section (Section 5).

2. Spatial disaggregation of information using a microdata approach and corrected with a Generalized Cross Entropy (GCE) method

Having good socioeconomic indicators at the local level is essential for analytical and research purposes as well as for policy design and evaluation. Nevertheless, statistical information is commonly available for large regions but only rarely available at more disaggregated levels or for smaller geographical units, such as LAU 2 regions in Europe. For this reason, there has been a broad development of techniques and methodologies to estimate disaggregated data. We can classify the alternative techniques to disaggregate spatial data into two broad classes: methodologies aimed at estimating values for the local spatial unit of analysis, *areal interpolation techniques*, and methodologies that take individual entities such as households and firms as the unit of analysis, *microdata-based techniques*.

2.1. A brief reappraisal of areal interpolation techniques for the spatial disaggregation of information

The first group of methodologies, *areal interpolation techniques*, is based on the idea of transforming data for a set of source zones into data for a set of target zones. The first attempts to estimate economic variables at a disaggregated scale can be linked to Tobler's (1979) spatial smoothing methodology, which has been improved to include more complex structures (see King et al., 2004, for an exhaustive review).

However, the disaggregation obtained using *areal interpolation techniques* is a graphical distribution of totals over a map. Several methodologies have tried to obtain disaggregated estimates while reducing inaccuracy problems; most of the estimates produced by such methods can be divided into direct and indirect estimates (see Rao, 2003). In the direct estimate methodologies, there are two options: 'model-based' estimators and 'design-based' estimators (see Pfeffermann, 2013, for a complete summary of the most recent developments in these two methodologies).

Model-based estimators try to extrapolate weights for each sub-area using an econometric model with other support variables that are related to the weight of the area. However, as shown in several articles (e.g., Hansen et al., 1983), these estimators face the important risk of high bias if the model is mis-specified.

Due to the risks involved with model-based estimates, researchers often turn to one of the best-known alternative estimators, the design-based methodology. This is a statistical approach that tries to obtain optimum sample weights to implement a sampling design. According to this methodology, small area estimations can be obtained when each sample is assigned an unbiased weighting. Some recent proposals within this methodology can be found in Jiang and Lahiri (2006) or Chandra and Chambers (2009), while a detailed review can be found in Rao (2003).

The main problem with direct estimators is that they usually result in wide confidence intervals due to problems of small sample size (even in the case of a correctly estimated model). These estimators also assume that it is possible to modify the design of the sampling process, and it is not obvious how to choose the weights. Hence, it has been necessary to devise methodologies that reduce these problems. The indirect method incorporates previously estimated information with out-of-sample data to adjust the estimations, thereby reducing the problems of variability in the estimations. A good example of this methodology can be found in Griffiths (1996).

A few *areal interpolation techniques* consider the special features of spatial data. Specifically, the spatial dependence effect can provide useful information in the spatial disaggregation procedure. Spatial dependence reflects a situation where values observed at one location depend on the observations at nearby locations. Benedetti and Palma (1994) introduced the Bayesian interpolation method (BIM), which exploits this general property of spatial data, to the areal interpolation problem. For recently proposed areal interpolation methods that consider the spatial nature of data, see Gotway et al. (2013) and Murakami and Tsutsumi (2011).

BIM requires some assumptions on the spatial data generating process. Commonly, spatially referenced data are considered to be a realization of a spatial stochastic process or random field, which is a collection of random variables indexed by their locations. With regard to the areal interpolation problem, data related to both the source and target zones can be interpreted as realizations of spatial stochastic processes. The spatial stochastic process generating the data related to the target zones (i.e., the areal units corresponding to the finer spatial scale) is referred to as the original process. The spatial stochastic process generating the data for the source zones (i.e., the areal units corresponding to the aggregated spatial level) is referred to as the aggregated process. Assuming that data are available only at the aggregated spatial level, the objective becomes to restore the realizations of the original process given the realization of the aggregated process.

The basic assumption on which BIM relies concerns the joint probability distribution of the original process, which is assumed to be a Gaussian distribution. The spatial dependence effect is taken into account by modelling the Gaussian random field by the conditional autoregressive (CAR) specification. This assumption does not entail any loss of generality since any Gaussian process on a finite set of sites can be modelled according to this specification. The CAR specification introduces the spatial dependence effect in the covariance structure of the process as a function of a scalar parameter of spatial autocorrelation and of a spatial weight matrix, which summarizes the proximity between any pairs of spatial units. Following a Bayesian approach, we can combine the prior information on the distribution of the original process with the data available at the aggregated spatial level to derive the posterior probability distribution of the original process. Benedetti and Palma (1994) derive the parameters of this posterior. Other examples of this approach, similar to those used to deal with a disaggregation problem, can be seen in Panzera et al. (2016), where missing data in spatial models are addressed through BIM.

In contrast to areal interpolation techniques, a second family of disaggregation techniques, known as microdata-based techniques, incorporates available information on individuals to estimate heterogeneity within regions. The additional evidence within regions helps to obtain more robust estimations than those produced by areal interpolation alone, reducing the confidence intervals and increasing the heterogeneity within groups.

2.2. A different approach to the spatial disaggregation of information: microdata-based techniques

Rather than to produce direct estimates of the variable of interest at the desired spatial scale, the basic idea of *microdata-based techniques* is to predict this variable at the level of individual agents (such as households, firms, workers, etc.). If the geographical location of these individual agents is observable at a highly disaggregated spatial scale, the indicators for small areas are calculated simply by summing or averaging the individual estimates. In most cases, however, researchers have to deal with microdata that do not provide the location of the individuals in a high level of detail. Surveys designed to study income distribution issues (such as household surveys) do not usually allow for a precise identification of the geographical location of the individuals surveyed. On the other hand, databases that do provide more precise spatial location data of the individuals, such as population census microdata, do not normally contain information on economic variables such as household income. The most important studies proposing a way to solve this problem are the contributions by Elbers et al. (2003) and Tarozzi and Deaton (2009), who proposed a modification to Elbers et al.'s (2003) method. The basic idea of both works consists of "projecting" predictions of the variable of interest onto the sample of households that form the population of a household survey.

We depart from a set of variables (x_h) that are observable with the same definition in two different databases and can explain the dependent variable. The idea of this first step is to estimate a parametric model using household data and then apply it to census data. Normally, household surveys do identify the aggregate region r of residence, so some heterogeneity in the process can be included. Therefore, the parameters in $\hat{\gamma}_r$ represent, for each region, the relation between exogenous variables in each household h and the dependent variable for each observable region r in household survey data. These parameters are estimated with parametric binary models (Probit) when we have a binary dependent variable.

The models estimated with household data are then applied to data on the same variables from households collected in census microdata to obtain the conditional probability, for binary variables, or expected value, for continuous ones. Therefore, the imputed poverty count of each local area i with a total of N_i households can be obtained as in equations **¡Error! No se encuentra el origen de la referencia.** and (2):

$$\hat{y}_i = \frac{1}{N_i} \sum_{h \epsilon H(i)} P(y_h \leq z | x_h; \hat{\gamma}_r) \tag{1}$$

$$\hat{y}_i = \frac{1}{N_i} \sum_{h \epsilon H(i)} E(y_h | x_h; \hat{\gamma}_r) \tag{2}$$

which is the mean of the conditional probabilities or expected values for the households in each local area defined in the census data. As in Tarozzi and Deaton (2009), the results of this estimation can be referred to as projection estimates.

2.3. Corrections to the estimations using the Generalized Cross Entropy (GCE) method

As explained by Tarozzi and Deaton (2009), the *microdata-based estimation technique* proposed in the previous section is subject to important uncertainty and errors that should be taken into consideration. In fact, projection estimates may lose accuracy from two main sources of bias: mis-specification of the proposed model and large variability in small sample areas.

In the first case, there is a plausible risk arising from the choice of the proper variables to explain the behaviour of the variable of interest. The researcher may omit relevant variables to predict the variable of interest, or data on such variables may not be available in the database. In this case, predictions may experience bias from the differences between the true theoretical model and the final specification behind the predictions.

Even if it is possible to specify the correct model, the difficulty of high variance in the predictions in areas with a small sample may remain. In this case, estimates may suffer from high variability, making them unreliable.

Estimates produced with these methods do not even have to be consistent with the aggregates that are already observable. Therefore, to reduce this uncertainty, we propose incorporating information on the observable aggregates to make the projection estimates consistent with it. We propose to make this correction combining Tarozzi and Deaton's (2009) method with the GCE estimator based on the work of Bernadini-Papalia and Fernández-Vázquez (2018).

The Bernadini-Papalia and Fernández-Vázquez (2018) estimator relies on additional information that is known and reliable outside a sample and adjusts the values of a variable to make them consistent with this information. This methodology is based on the framework of maximum entropy. In this framework, the variable of interest has a probability distribution with an unknown probability for each value. The basic idea of this type of methodology is to obtain the estimation with the highest degree of uncertainty that at the same time is able to fulfil the conditions of the observable data. Therefore, the set of probabilities has to be calculated through optimization of an entropy function, as in equation **¡Error! No se encuentra el origen de la referencia.** (see Shannon (1948) for additional details):

$$Ent(p) = -\sum_{m=1}^{M} p_m \ln(p_m) \tag{3}$$

This function has its maximum value for a distribution of probabilities p_m of a variable of interest that represents the homogeneous distribution. Therefore, the optimization process tends to reduce the minimum distance possible that the restrictions allow. Any new information about the variable of interest would restrict the feasible region of the optimization, moving the optimum value away from the homogeneous distribution.

With this same idea, the Generalized Maximum Entropy (GME) estimator (see Golan et al., 1996) considers the coefficients of a linear regression b_{km} as discrete random variables with M different possible values for each k coefficient. The optimization problem calculates the probability of the different possible values of the regression coefficients. Once the vector of probabilities for each coefficient p_{km} has been calculated, the estimated set of coefficients $\tilde{\beta}_k$ can be calculated as its expected value – given the estimated probabilities for each value of their discrete distribution; see equation **¡Error! No se encuentra el origen de la referencia.**:

$$\tilde{\beta}_k = \sum_{m=1}^{M} p_{km} b_{km} \ ; \ k = 1, \dots, K \tag{4}$$

GME redesigns the linear regression estimation as an optimization problem of the entropy function in equation **¡Error! No se encuentra el origen de la referencia.** subject to the linear relation between the exogenous variables and the independent variables. This methodology does not require a restrictive assumption for the error. It only requires a finite matrix of variance and covariances as well as a null expected value in the errors. With these assumptions, the errors are also presented as a discrete random variable with J possible values with a set of probabilities u_j for the possible value of the error presented in v_j. The final optimization problem in GME for a cross-sectional database is summarized in equations **¡Error! No se encuentra el origen de la referencia.**, **¡Error! No se encuentra el origen de la referencia.**, **¡Error! No se encuentra el origen de la referencia.** and **¡Error! No se encuentra el origen de la referencia.**:

$$\underset{\boldsymbol{P},\boldsymbol{U}}{Max}\, Ent(\boldsymbol{P},\boldsymbol{U}) = -\sum_{k=1}^{K}\sum_{m=1}^{M} p_{km}\, ln(p_{km}) - \sum_{j=1}^{J} u_j\, ln(u_j) \tag{5}$$

subject to:

$$\boldsymbol{y_r} = \sum_{k=1}^{K}\sum_{m=1}^{M} \boldsymbol{b_{km}} p_{km}\, \boldsymbol{\widehat{x}_h} + \sum_{j=1}^{J} v_j\, u_j \tag{6}$$

$$\sum_{m=1}^{M} p_{km} = 1\,;\ k = 1, \dots, K \tag{7}$$

$$\sum_{j=1}^{J} u_j = 1 \tag{8}$$

Through this optimization problem, GME ensures that the optimal point is also compatible with the feasible region defined by the linear relation between the observed values in the dependent and exogenous variables.

We propose to apply the modification of GME defined in Bernadini-Papalia and Fernández-Vázquez (2018) as a consistent method to update projection estimates in equations **¡Error! No se encuentra el origen de la referencia.** and **¡Error! No se encuentra el origen de la referencia.**. The variable of interest in this estimator is a disaggregated variable or estimation that is not consistent with a known aggregate. This variable can be expressed as in equation **¡Error! No se encuentra el origen de la referencia.** thanks to a support vector like the one of equation **¡Error! No se encuentra el origen de la referencia.**:

$$\tilde{y}_i = \sum_{m=1}^{M} p_{im} b_{im} \tag{9}$$

The 'direct GME' approach in this methodology proposes a feasible region where the weighted mean of i observations at a lower level – with population size equal to N_i – has to be equal to an out-of-sample aggregated value $\bar{y}$ – see equation **¡Error! No se encuentra el origen de la referencia.**. As in GME, this model includes an error with similar assumptions, but the notation is adapted to the problem. In this case, the errors are also a discrete random variable with L possible values with a set of probabilities w_{il} for the possible value of the error presented in v_{il}. However, the restriction has been modified, given that the weighted mean of the variable of interest has to be equal to the observable aggregate:

$$\underset{\boldsymbol{P},\boldsymbol{W}}{Max}\, Ent(\boldsymbol{P},\boldsymbol{W}) = -\sum_{i=1}^{D}\sum_{m=1}^{M} p_{im}\, ln(p_{im}) - \sum_{i=1}^{D}\sum_{l=1}^{L} w_{il}\, ln(w_{il}) \tag{10}$$

subject to:

$$\bar{\boldsymbol{y}} = \sum_{i=1}^{D}\left[\sum_{m=1}^{M} \boldsymbol{p_{im}} b_{im} + \sum_{l=1}^{L} w_{il}\, v_{il}\right]\left[{N_i}\Big/{\sum_{i=1}^{D} N_i}\right] \tag{11}$$

$$\sum_{m=1}^{M} p_{im} = \sum_{l=1}^{L} w_{il} = 1\,;\ i = 1, \dots, D \tag{12}$$

The application of this estimator to the disaggregation framework of Elbers et al. (2003) or Tarozzi and Deaton (2009) enhances the estimations with information about the aggregates, but the estimator is also a flexible tool. For example, in some cases, statistical institutes do not allow researchers to work with microdata but this tool could be used to update estimations at an individual level or, when this is not possible, at a local level.

3. An empirical application: Local poverty estimations for France, Portugal, Spain and the United Kingdom

In this section, we estimate poverty indicators at the local level using the methodology proposed in the previous section. Specifically, we disaggregate the regional AROPE index (the propensity to be at risk of poverty and exclusion), available at the regional level thanks to the EU-SILC (European Union Survey on Income and Live Conditions), for four EU countries (Spain, France, the United Kingdom and Portugal).

The AROPE index or rate is the measure of poverty officially used in the EU. It is a multidimensional index that combines income and non-income indicators and considers individuals to be at risk of poverty if they fulfil at least one of three criteria: (i) they have a disposable income below 60% of the national mean; (ii) they face severe material deprivation, decided based on a set of household consumption goods and services[2]; and/or (iii) they live in a household with a low work intensity.

[2] The nine basic consumption concepts used are (1) late payment of rent, mortgage or utility bills of the primary residence over the last 12 months; (2) inability to keep the home adequately heated; (3) inability to take at least a one-week holiday each year; (4) inability to eat meat or protein at least every two days; (5) insufficient money for unforeseen expenses; (6) no telephone; (7) no colour television; (8) no washing machine; and (9) no car.

The AROPE rate is reported at the country level, and for most EU countries, it is also reported at either the NUTS 1 or NUTS 2 level (with Germany and the United Kingdom before 2010 being notable exceptions). The contribution of this study is providing AROPE figures at a smaller spatial scale than the NUTS 1 and NUTS2 regions, which will enable better identification of the localities within the geographically broad NUTS regions where tackling poverty is most necessary.

3.1. Dataset, variables and country selection

The methodology described in the previous section requires the combination of two datasets: the EU-SILC (provided by Eurostat and population census microdata) and national population census microdata, which are only provided by the national institutes of statistics of each country, in some cases under strict restrictions. As population censuses are only released every 10 years and the last year available is 2011, the local estimations obtained in this empirical application correspond to 2011.

The data from the EU-SILC only provide information on the NUTS 2 or NUTS 1 region (depending on the country) where the respondent resides, so in practical terms, it is not possible to perform any socioeconomic analysis within the region. However, information from the EU-SILC can be “matched” with each national population census microdata, a much larger sample with information on the locality (in some cases LAU 2 regions) where the individuals reside. Despite having such an advantage against the EU-SILC, these microdata provide no information on income, poverty or wellbeing figures, precluding any analysis of income or poverty inequality within regions. Table 1 summarizes the basic characteristics and sample sizes for both the EU-SILC and population censuses in the EU countries under analysis.

Table 1. EU-SILC and census sample size and location level.

	France	Portugal	Spain	United Kingdom
Households in SILC	11,360	5,740	13,109	8,058
SILC-Location level	NUTS2	-	NUTS2	NUTS1
Households in census	8,741,050	204,409	1,619,806	1,312,291
Census location level	Canton	NUTS3 + 5 cities	Municipalities	Local areas

The population censuses provide information on the place of residence of individuals/households at a smaller spatial scale than that of the EU-SILC. However, the local spatial units available in each census differ. Thus, *cantons* for France and *municipalities* for Spain correspond to the local administrative units (LAUs) suggested by Eurostat. The spatial partition used by the United Kingdom in its 2011 population census is very similar to the LAU official division, but a few *districts* or *individual unitary authorities* have been merged for confidentiality reasons. However, in Portugal, unless the individuals reside in the 5 largest cities, only information at the NUTS 3 level is provided, which might add complexity to any attempt to compare these countries.

The variables chosen to disaggregate the variable of interest, i.e., the AROPE index, are similar to those in Tarozzi and Deaton (2009), and we include all possible relevant variables from the censuses that are also in the EU-SILC and share a common definition. To make the methodology more consistent between countries, the variables have to be chosen considered that there are cases in which some information was discarded because not all countries provide a similar variable or concept.

The set of exogenous variables can be divided into two groups. In the first group, the variables are related to the head of household. In the second group, the variables provide information on the characteristics of the whole household. The household head is defined – in order – by employment status[3], position in the employment hierarchy, level of education, age and gender (to account for possible labour discrimination). UK Office of National Statistics provides their own identification of the household reference person according to labour status. In this case, the EU-SILC identification of the household head follows the same criteria.

The variables chosen for each country are shown in Tables 2 and 3. Despite the differences between national censuses, the set of variables for all countries is almost the same – labour activity and personal characteristics of the head of household and structure of the household. However, minor differences might be found in each case. To accommodate these differences, the GME estimation with the EU-SILC in each country is made with the information available in the corresponding census.

3.2. Main results

The application of the methodology proposed in Section 2 to the data described in the previous section allows us to obtain disaggregated estimates of the AROPE index. We should recall that the omission of relevant variables to predict the variable of interest because of database limitations could bias the predictions due to differences between the true theoretical model and the final specification. The main results are summarized in Figure 1. In Figure 1.B, the local information estimated is represented and can be compared with the official Eurostat regional figures using the EU-SILC represented in Figure 1.A.

This first general map allows us to see that intraregional spatial differences are certainly very relevant. As the new economic geography approach predicts, the centre-periphery duality is not only observable at the national and international scales but is also happening within regions on a local scale. There are strong regional disparities in the countries under analysis[4], with low AROPE rates for central regions, i.e., regions that contain the capital city of the country or large metropolitan/urban areas, and higher rates for peripheral regions. We will use the next section to offer more detailed comments on the observable spatial patterns.

[3] In the EU-SILC, the main job classifications for this variable are self-employed with employees, self-employed with no employees, employee and family worker.

[4] Regional disparities in Portugal cannot be shown, as the EU-SILC database does not provide information at NUTS levels.

Table 2. Information from the head of the household used as predictors.

France	Portugal	Spain	United Kingdom
Age and age^2	*Age and age*^{2}	*Age and age*^{2}	*Age and age*^{2}
Gender	*Gender*	*Gender*	*Gender*
Immigrant	*Foreigner:* -EU country -Non EU country	*Foreigner:* -EU country -Non EU country	*Foreigner:* -EU country -Non EU country
Marital status: -Married -Widow -Divorced	*Marital status:* -Married -Separated -Widow -Divorced	*Marital status:* -Married -Separated -Widow -Divorced	*Marital status:* -Married -Separated -Widow -Divorced
Education: -Post-mandatory non-college -College	*Education:* -Post-mandatory non-college -College	*Education:* -Post-mandatory non-college -College	*Education:* -Post-mandatory non-college -College
Activity status: -Worker -Retired or disabled -Other activity	Activity status: -Worker -Retired or disabled -Other activity	Activity status: -Worker -Retired or disabled -Other activity	Activity status: -Worker -Retired or disabled -Other activity
Partial employment	*Partial employment*	*Partial employment*	*Partial employment*
	Occupation: -Manager -Technician or professional -Support worker or sales -Craft, machine operators or skilled agricultural worker	*Occupation:* -Manager -Technician or professional -Support worker or sales -Craft, machine operators or skilled agricultural worker	*Occupation:* -Manager -Technician or professional -Support worker or sales -Craft, machine operators or skilled agricultural worker
Economic sector: -B, C, D, or E -F -G -H -I -J -K -L, M or N -O, P or Q -R, S, T or U	*Economic sector:* -F -G -H -I -J -K -L, M or N -O -P -Q -R, S, T or U	*Economic sector:* -C,D or E -F or L -G -H or I -K, J, M or N -O -P -Q -R, S, T or U	*Economic sector:* -F -G -H -I -L, M or N -O -P -Q -R, S, T or U

Note. Sectors are defined according to the statistical classification of economic activities in the European Community (NACE).

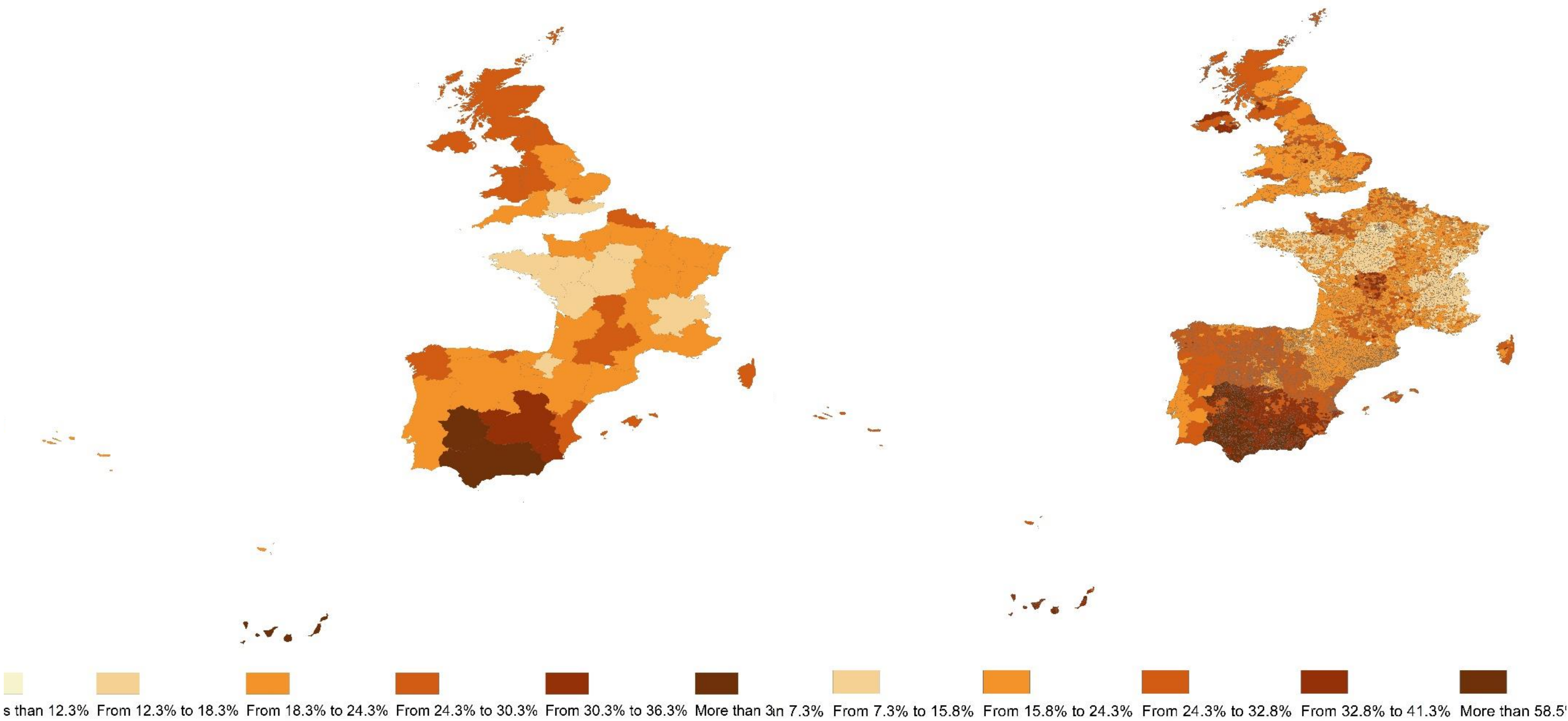
Table 3. Characteristics of the household used as predictors.

France	Portugal	Spain	United Kingdom
Tenure: -Tenant -Free accommodation	*Tenure:* -Tenant	*Tenure:* -Owner with mortgage -Tenant -Free accommodation	*Tenure:* -Owner with mortgage -Tenant or free accommodation
Number of rooms		*Number of rooms*	
Family structure: -Single parent household -Couple with no dependent children -Couple with dependent children -Other family	*Family structure*: -Single parent household -Couple with no dependent children -Couple with dependent children	*Family structure:* -Single parent household -Couple with no dependent children -Couple with dependent children -Other family with no dependent children -Other family with dependent children	*Family structure:* -Single parent household -Couple -Other family
Workers in the household	*Workers in the household*	*Workers in the household*	*Members in the household*
Members between 0-4 years Members between 5-15 years Members between 16-24 years Members between 25-34 years Members between 35-64 years Members between 65-84 years Members older than 85 years	Members between 0-4 years Members between 5-15 years Members between 16-24 years Members between 25-34 years Members between 35-64 years Members older than 65 years	Members between 0-4 years Members between 5-15 years Members between 16-24 years Members between 25-34 years Members between 35-64 years Members between 65-84 years Members older than 85 years	

Figure 1. AROPE index (or at risk of poverty and exclusion rate) for France, Portugal, Spain, and United Kingdom (2011).

A. At regional level with EU-SILC

B. At local level with own estimations



Source: EU-SILC, Eurostat (2011), Census and own estimations.

4. Discussion of the results: the relevance of intraregional inequalities and metropolitan disparities

Within the EU overall, the highest AROPE values tend to be appear mainly for eastern European regions and southern regions, Portugal, southern Spain and southern Italy (Fernandez-Vazquez et al. 2019). When we look only at the 2011 regional AROPE rates for our four EU countries under study (Figure 1. A), relatively high values can be found for Spain, Portugal and the peripheral regions of France and the UK. Particularly high values are present in the peripheral Spanish regions of the south.

It is important to be aware that the poverty estimates shown are for *localities*, and low/high values mean that the proportion of households at risk of poverty is relatively low/high compared with that of other localities across the EU. Previous studies confirm the existence of strong interregional poverty inequalities at the EU level, with *EU capital regions*[5] commonly showing the best performance in AROPE rates (low values) and on many other indexes (Athanasoglou and Dijkstra, 2014). This "capital status regularity" can be confirmed for the countries under analysis by looking at Figure 1.A, where the NUTS 2 regions of Madrid, Paris and London show lower percentages than the country averages (although not the lowest country values). Nevertheless, the spatial limitations of the official data available (AROPE figures at the NUTS 2 level) mask the internal heterogeneity that exists in both capital and non-capital regions, as shown by Figure 1.B presenting our AROPE estimations at the local level.

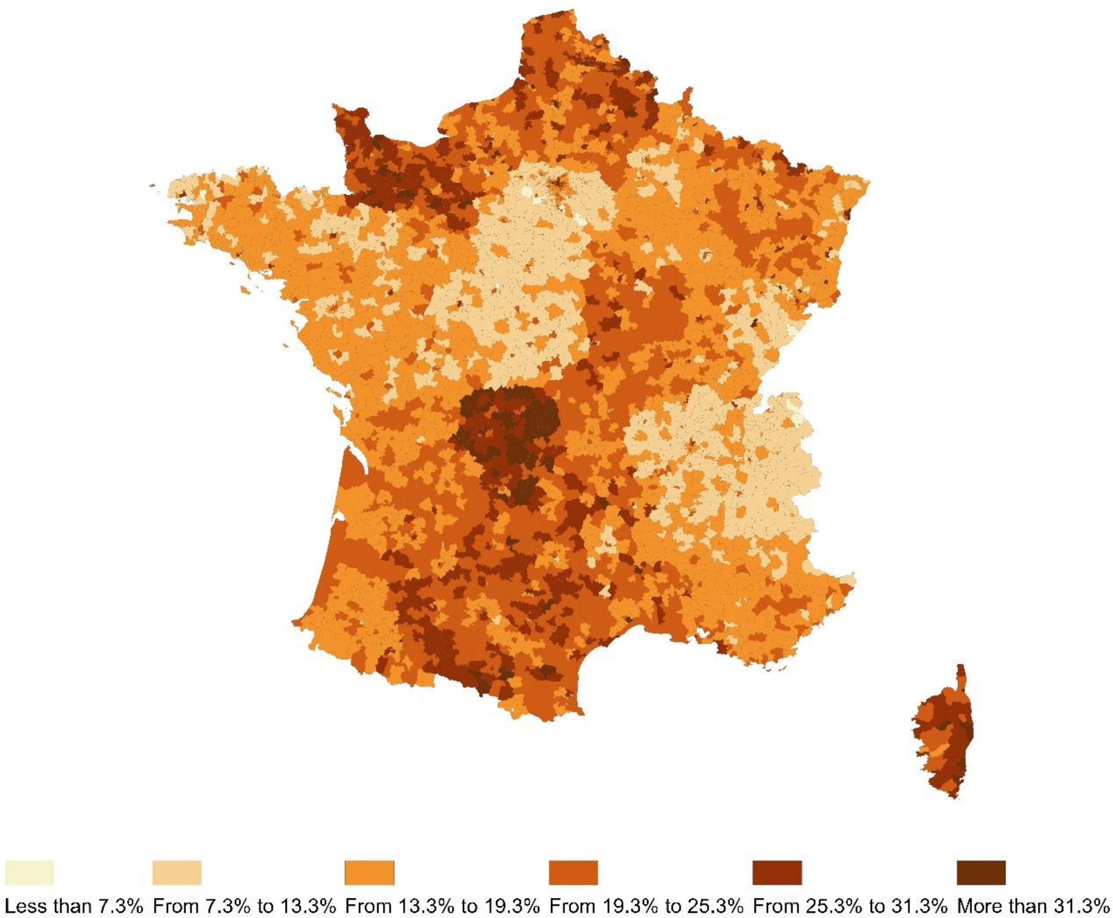
Figures 2 to 5 show local AROPE estimates obtained for each individual country. These figures allow us to appreciate in greater detail the existing internal regional disparities regarding poverty. Local AROPE figures are mapped using each national AROPE rate average as a benchmark. The main dynamics already observable in the map of large regions continue in the disaggregated representations. There is a centre/periphery dynamic on a continental as well as national scale. The central axis that connects the main metropolises of Europe – London (in the pre-Brexit EU), Paris, Amsterdam, Hamburg and Vienna, among others – shows a concentration of the lower levels of risk of poverty and social exclusion. When we move away from this central axis to the south or east or north, we can see a rise in the poverty incidence. Inside each country, we also observe how the closer an area is to this European central axis, the lower on average is the incidence of the risk of poverty and social exclusion.

The two main hypotheses that implicitly motivate this exercise of spatial disaggregation of poverty figures using the AROPE concept are confirmed just by observing the new maps generated. First, there is significant intraregional heterogeneity in the poverty figures in the four studied countries. Second, spatial inequalities inside the local context as well as patterns of concentration of poverty are observable.

[5] Brussels is a renowned exception to this regularity.

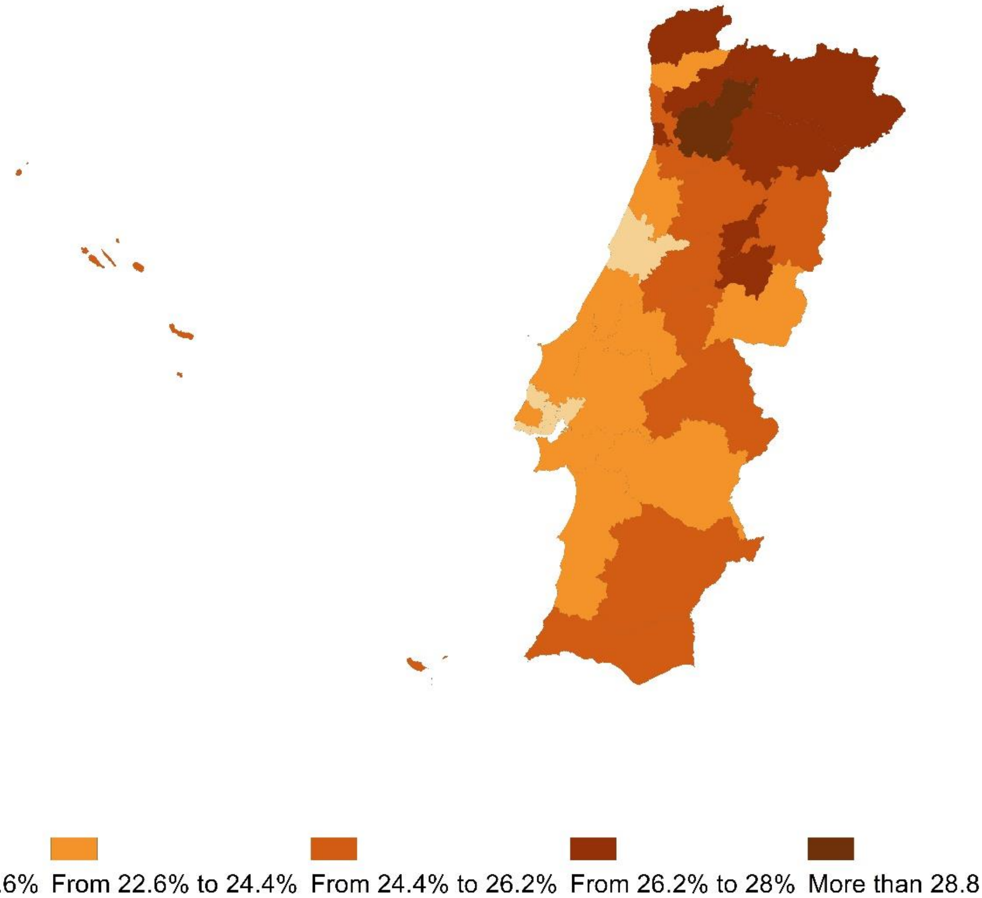
In the disaggregated maps, the intraregional heterogeneity is observable, with pockets of rich areas surrounded by poor towns and vice versa. This happens in the four considered countries but is especially clear in France, Spain and the United Kingdom thanks to the higher level of disaggregation of the local units in these three cases. For instance, the south of Spain has a significantly higher risk of poverty incidence, but there are specific municipalities with levels above the European average. This also occurs in other areas, such as the south of France or the north and east of the United Kingdom. When we analyse the intraregional behaviour, it is observable, in general, that the centre-periphery pattern is present.

Figure 2. AROPE rate estimations at local level (French cantons): 2011



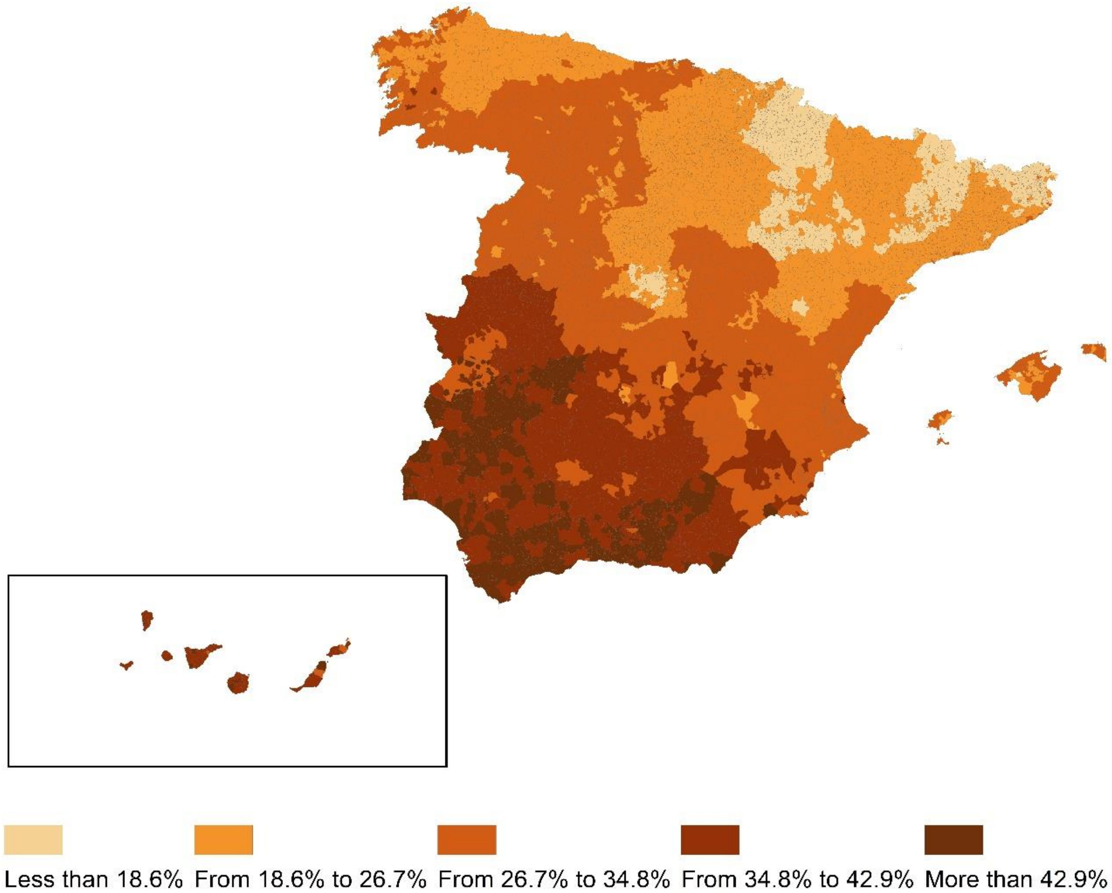
Source: Own estimations.

Figure 3. AROPE rate estimations at local level (Portuguese NUTS3 regions and 5 cities): 2011



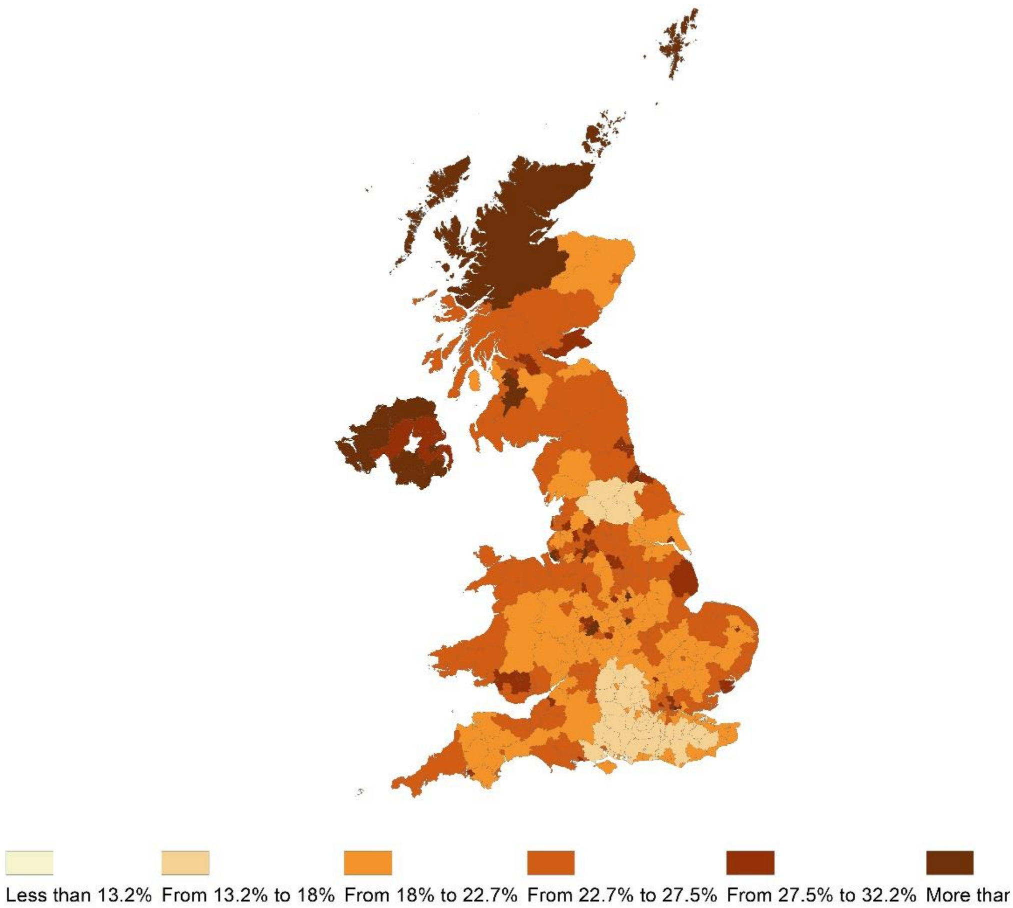
Source: Own estimations.

Figure 4. AROPE rate estimations at local level (Spanish municipalities): 2011



Source: Own estimations.

Figure 5. AROPE rate estimations at local level (United Kingdom): 2011



Source: Own estimations.

Prominent spatial inequalities at the local level can be detected in the capital regions, with large value gaps between the core city and the localities around the core that together comprise the metropolitan area or the greater city. Thus, the high poverty values and poor performance in terms of other relevant indicators outlined by Athanasoglou and Dijkstra, L. (2014) for the NUTS 2 region of Brussels (BE10) is not an exception but a result of the comparatively different spatial scales used to define the NUTS 2 regions of Madrid, Paris, London and Lisbon. While the NUTS 2 region of Brussels comprises almost exclusively the city of Brussels, the NUTS 2 regions of Paris, Madrid, London and, to a lesser extent, Lisbon comprise geographically much larger areas, with one core city, its surrounding neighbouring towns/cities and many rural localities[6].

[6] For a discussion on the boundaries of a city, a larger urban zone (or LUZ) and a greater city, see the *Methodological manual of territorial typologies* (Eurostat, 2018).

Indeed, the AROPE disaggregated figures for the above NUTS 2 capital regions show the relevance of the spatial unit of analysis chosen and the existing inequalities within capital regions (Figure 6)[7]. The city of Madrid (as opposed to the Madrid NUTS 2 region) shows low AROPE figures in comparison to those of the rest of the surrounding localities, with a clear centre/periphery pattern consistent with the distribution of income in the dense and compact European traditional cities. The region of Madrid is a clear example of this, with a dense, clearly identified city centre (Madrid city, Las Rozas, Pozuelo, Majalahonda, Boadilla del Monte, Tres Cantos) that exhibits low AROPE rates and AROPE values increasing as distance from the centre increases. For the case of London, a north-south pattern in the spatial distribution of poverty can be observed, with high AROPE rates in the north-eastern areas (Newham and Brent) and low values in the residential and southern peripheral areas (Richmond upon Thames, Sutton and Bromley). In the region of Paris, the city centre shows mean AROPE rates, while high AROPE rates are all clustered in the northern outer communes of La Courneve, Aubervillies, Garges les Gonese and Sainte Denis. However, while many cities from both southern and northern European countries (such as Stockholm and Milan) have a rich centre and poorer peripheries, others (such as Brussels and many eastern cities) are characterized by an urban centre occupied by the poor. As Cassiers and Lesteloot (2012) conclude, “the spatial lay-outs of inequalities in European cities are much more complex and diversified than the rude distinction between rich and poor centres and peripheries” (pp. 1919).

Recognizing the existence of local features and quantifying the spatial heterogeneity within regions and cities is crucial for the success or failure of any cohesive social policy (Minglione, 2004). While social cohesion has been mainly funded at the regional level, it should be analysed and diagnosed at the local level, at the level of cities as a whole or even at a smaller scale.

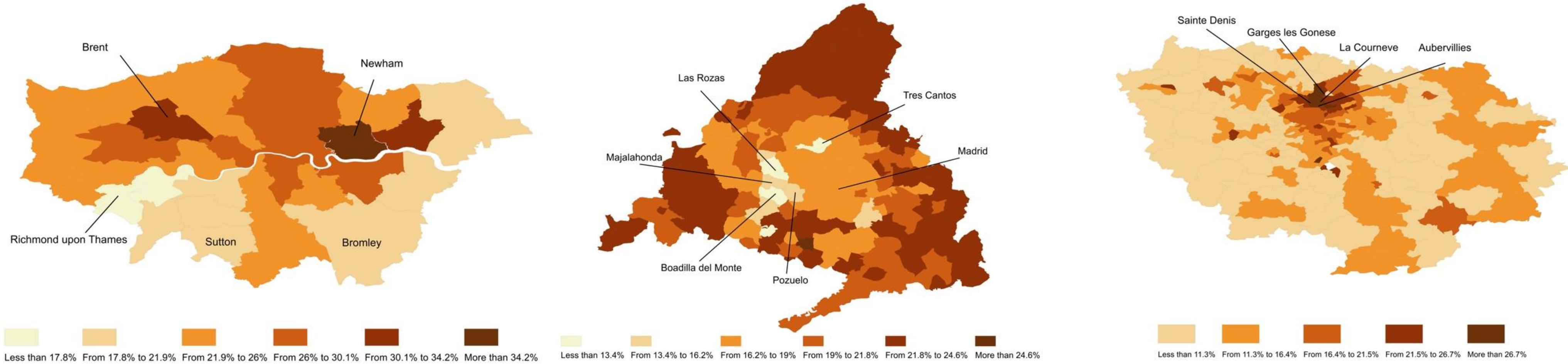
[7] Lisbon map not shown.

Figure 6. AROPE rate within NUTS2 regions of London, Madrid and Paris (2011).

A. London (Greater London)

B. Madrid (Autonomous Community of Madrid)

C. Paris (Metropolitan Area of Paris)



Source: EU-SILC, Eurostat (2011), Census and own estimations.

5. Conclusions

In the EU, territorial inequalities in terms of income or poverty have been broadly analysed at the national or regional level, especially after the spatially uneven impact of the global economic crisis was confirmed. However, mainly due to the lack of reliable data, very little attention has been paid to territorial inequalities within European regions, i.e., at a more local level, such as in metropolitan areas, cities or neighbourhoods.

This paper proposes a combination of the Tarozzi and Deaton (2009) and Bernadini-Papalia and Fernández-Vázquez (2018) methodologies to disaggregate the official regional poverty figures contained in EU household surveys into poverty indicators for smaller spatial units, mainly local administrative units or LAU 2. The household survey poverty figures at the regional level are combined with microcensus data for each country that contain details on the local entities of residence to disaggregate the regional poverty indicator.

While similar disaggregation methods have been applied by the World Bank to map poverty (and income) inequality at the local level in countries such as Cambodia, Mexico, Morocco, South Africa and Uganda, no comprehensive procedures have been applied before at the European local level. This paper proposes a second step that guarantees consistency between the local poverty estimates and the regional poverty figures by adjusting the initial estimates based on generalized cross entropy (GCE), extending previous methodologies.

The population at risk of poverty indicator, which is the poverty indicator reported in the EU-SILC at the regional level, is spatially disaggregated into local administrative units for four European countries for which microcensus databases were accessed: France, Spain, the United Kingdom and Portugal. In the disaggregated maps, the intraregional heterogeneity is clearly visible: rich pockets surrounded by poor towns and vice versa. When we specifically focus the attention in the intraregional maps, in general, a reproduction of the centre/periphery laws that operate at national and continental levels is observable. Additionally, prominent spatial inequalities at the local level can be detected in the capital regions, with large value gaps between the core city and the localities around the core that together comprise the metropolitan area or the greater city.

This work is a first contribution in the direction of better understanding the intensity and patterns of economic inequalities at the local scale. Although it is important to understand the differences between regions and countries, increasingly intense socioeconomic gaps are being generated within regions or even within cities. Our results fully confirm the importance of these local inequalities and the need to articulate highly spatially disaggregated economic cohesion policies. In this sense, European regional policy is acting on a scale, the regional one, which presents high internal heterogeneity. The selection of intervention areas and the design of actions should take place at the local level based on a more detailed understanding of intraregional inequalities.

6. References

Athanasoglou, S. and Dijkstra, L. (2014) *The Europe 2020 Regional Index (Report EUR 26713)*. European Commission. Joint Research Center.

Ballas, D.; Dorling, D. and Hennig B. (2017) Analyzing the regional geography of poverty, austerity and inequality in Europe: a human cartographic perspective, *Regional Studies*, 51(1), 174-185.

Benedetti, R. and Palma, D. (1994) Markov random field-based image subsampling method summary, *Journal of Applied Statistics*, 21(5), 495–509.

Bernadini-Papalia, R. and Fernández-Vázquez, E., (2018) Information theoretic methods in small domain estimation, *Econometric Reviews*, 37(4), 347–359.

Cassiers, T. and Kesteloot, C. (2012) Socio-spatial Inequalities and Social Cohesion in European Cities, *Urban Studies*, 49 (9), 1909-1924.

Chandra, H. and Chambers, R. (2009) Multipurpose Small Area Estimation, *Journal of Offic*ial Statistics, 25, 379–395.

Copus, A.; Melo, PC.; Kaup, S.; Tagai, G. and Artelaris, P. (2015) Regional poverty mapping in Europe – Challenges, advances, benefits and limitations. Local Economy 0(0), 0. 1- DOI: 10.1177/0269094215601958)

Darvas, Z. (2016) Some Are More Equal Than Others: New Estimates of Global and Regional Inequality, *Bruegel Working Paper 8*.

Dauderstädt, Michael, and Cem Keltek (2015) *Social Europe in the crisis*, Berlin, Friedrich-Ebert-Stiftung.

Davies, J.B.; Lluberas, R. and Shorrocks, A. (2017) Estimating the Level and Distribution of Global Wealth, 2000-2014, *Review of Income and Wealth*, 63 (4), 731-759.

Dijkstra, L.; Poelman, H. and Rodriguez-Pose, A. (2019) The geography of EU discontent, *Regional Studies* (on line).

Duro, J.A. (2004) Regional income inequalities in Europe: an updated measurement and some decomposition results, *Facultat de Ciències Econòmiques i Empresarials: Working Paper No. 4.11*.

Elbers, C.; Lanjouw, J.O. and Lanjouw, P. (2003) Micro-Level Estimation of Poverty and Inequality. *Econometrica*, 71(1), 355–364.

Eurostat (2018) *Methodological manual on territorial typologies.* Eurostat.

Fernandez-Vazquez, E; Plotnikova, M.; Postiglione, P.; Rubiera-Morollon, F. and Viñuela, A. (2018) *Deliverable 2.1 Review of Official Data*. IMAJINE WP2 Analysis of Territorial Inequalities in Europe.

Fratesi, U. and Rodríguez-Pose, A. (2016) The crisis and regional employment in Europe: what role for sheltered economies?, C*ambridge Journal of Regions, Economy and Society*, 9(1), 33-57.

Fujita, M. and Krugman, P. (1995) When is the economy monocentric?: von Thünen and Chamberlin unified. *Regional Science and Urban Economics*, 25(4), 505–528.

Fujita, M.; Krugman, P. and Venables, A., (2001) *The spatial economy: Cities, regions, and international trade*, MIT press.

Gehlke, C.E. and Biehl, K. (1934) Certain effects of grouping upon the size of the correlation coefficient in census tract material, *Journal of the American Statistical Association*, 29(185), 169–170.

Golan, A.; Judge, G. and Miller, D. (1996) *A maximum Entropy Econometrics: Robust Estimation with limited data*, New York: Wiley.

Gotway, C.; Young, L.J.; Mar, N., and Young, J. (2013) A Geostatistical Approach to Linking Geographically Aggregated Data from Different Sources, *Journal of Computational and Graphical Statistics*, 16(1), 115–135.

Griffiths, R. (1996) Current population survey small area estimation for congressional districts, *Proceedings of Section on Survey Research Methods, American Statistical Association*, 314–319.

Hansen, M.H.; Madow, W.G. and Tepping, B.J. (1983) An evaluation of model-dependent and probability-sampling inferences in sample surveys, *Journal of the American Statistical Association*, 78(384), 776–793.

Heidenreich, M. and Wunder, C. (2008) Patterns of Regional Inequality in the Enlarged Europe, *European Sociological Review,* 24 (1), 19-36

Jiang, J., and Lahiri, P. (2006) Estimation of finite population domain means: A model-assisted empirical best prediction approach, *Journal of the American Statistical Association*, 101(473), 301–311.

King, G.; Rosen, O. and Tanner, M.A. (2004) Ecological Inference: New Methodological Strategies, *Analytical Methods for Social Research*. Cambridge University Press.

Krugman, P. (1991) Increasing Returns and Economic Geography. *The Journal of Political Economy*, 99(3), 483–499.

Krugman, P. and Venables, A.J. (1995) Globalization and the Inequality of Nations. *The Quarterly Journal of Economics*, 110(4), 857–880.

Milanovic, B. (2002) True world income distribution, 1988 and 1993: first calculation based on household surveys alone, *Economic Journal*, 112, 51-92.

Milanovic, B. (2005) *Worlds Apart: Measuring International and Global Inequality*. NPrinceton University Press. New Jersey.

Miranti, R.; McNamara, J.; Tanton, R. and Harding, A. (2011) Poverty at the Local Level: National and Small Area Poverty Estimates by Family Type for Australia in 2006, *Applied Spatial Analysis and Policy*, 4 (3), 145–171.

Murakami, D. and Tsutsumi, M. (2011) A new areal interpolation method based on spatial stadistics, P*rocedia - Social and Behavioral Sciences*, 21, 230–239.

Openshaw, S. (1984) *The modifiable areal unit problem, Concepts and techniques in modern geography*, GeoBooks.

Openshaw, S., Taylor, P., (1979) A million or so correlation coefficients. *Statistical Methods in the Spatial Sciences*, 127–144.

Panzera, D., Benedetti, R., Postiglione, P., (2016) A Bayesian Approach to

Parameter Estimation in the Presence of Spatial Missing Data. *Spatial Economic Analysis*, 11(2), 201–218.

Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*, 28(1), 40–68.

Piketty, T. (2014) *Capital in the Twenty-First Century*. Harvard University Press. Cambridge.

Puga, D. (2002) European regional policies in light of recent location theories, *Journal of Economic Geography*, 2, 373-406.

Ramos, R. and Royuela, V. (2014) Income inequality in Europe. Analysis of recent trends at the regional level. W*orking Paper 2014/13 1/53*. Research Institute of Applied Economics, Regional Quantitative Analysis Research Group. Universitat de Barcelona.

Rao, J.N.K. (2003) S*mall Area Estimation*. Wiley. New Jersey.

Shannon, C.E., (1948) A mathematical theory of communications. *Bell System Technical Journal*, 27, 379–423.

Storesletten, K.; Telmer, C.I. and Yaron, A. (2004) Cyclical Dynamics in Idiosyncratic Labor Market Risk, *Journal of Political Economy,* 112, 695–717.

Tarozzi, A. and Deaton, A., (2009) Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas, *The Review of Economics and Statistics*, 91(4), 773–792.

Tobler, W.R. (1979) Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*, 74(367), 519–530.

Vacas-Soriano, C. and Fernández-Macías, E. (2018) I*ncome Inequality in the Great Recession from an EU-wide Perspective*, CESifo Forum, ISSN 2190-717X, ifo Institut - Leibniz-Institut für Wirtschaftsforschung an der Universität München, München, 19(2), 9-18.

Węziak-Białowolska, D. (2015) Poverty in the Regions of the European Union – Measurement with a Composite Indicator, C*ontemporary Economics,* 9(2), 113-154.