Working Paper Series

Measures of poverty and inequality in the countries and regions of EU

Nicholas T. Longford
Maria Grazia Pittau
Roberto Zelli
Riccardo Massari

ECINEQ WP 2010 – 182

ECINEQ 2010-182
October 2010

www.ecineq.org

Measures of poverty and inequality in the countries and regions of EU

Nicholas T. Longford
SNTL and UPF, Barcelona

Maria Grazia Pittau*, **Roberto Zelli, and Riccardo Massari**
Sapienza University, Rome

Abstract

The European Union Survey on Income and Living Conditions (EU-SILC) is the main source of information about living standards and poverty in the member states of the European Union. It provides reliable statistics at national level but sample sizes do not allow reliable estimates at sub-national level, despite a rising demand from policy makers and local authorities. We provide a comprehensive map of median income, inequality (Gini coefficient and Lorenz curve) and poverty (poverty rates), at country and regional levels, based on the equivalized household income in all the countries in which EU-SILC is conducted. We focus on personal income distribution within regions as opposed to per capita income distribution across regions to give a deeper insight into regional disparities. Small-area estimation is applied to improve estimates in regions with small sample size. Uncertainty of such complex non-linear statistics is assessed by bootstrap methods. Household-level sampling weights are taken into account in both the estimates and their relative bootstrapped standard errors.

Keywords: European regional economics measurement; EU-SILC; Gini coefficient; Poverty rates; Small-area estimation.
JEL classification: D31, I32, R10, C13.

* **Address of correspondence:** M. Grazia Pittau, Department of Statistics, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome; email: grazia.pittau@uniroma1.it.

1 Introduction

Reduction of regional disparities in the European Union (EU) has become an important goal after its latest enlargements in 2004 and 2007. The accession of 12 new members, each with less than half of GDP per capita of the earlier members of EU, has brought new challenges for a reinforced cohesion policy. Such a policy takes as its inspiration Article 130a of the Treaty on European Union and intends to "reduce disparities between the levels of development of the various regions and the backwardness of the least favoured regions, including rural areas" (First Report on Economic and Social Cohesion, 1996). At present, more than of two-thirds of the Structural Fund budget is allocated to regions in which the GDP per capita lags behind the EU average. Regional disparities have been extensively investigated in the economic literature by considering GDP per capita as a measure of disparity.

The effectiveness of interventionist regional policies has been often evaluated in terms of convergence or divergence of per capita income in the regions, eventually giving to each region a weight proportional to its population size (see, among others, Barro, 1991; Quah, 1996; Le Gallo, 2004; Pittau and Zelli, 2006). Findings depend on the time span examined, the number of regions, the level of disaggregation and the statistical method used. There is a widespread agreement that income disparities across European regions belonging to the EU15[1] have narrowed over time, but reduction of income disparities across regions cannot be equated with reduction of disparities within regions. That is, a region with high GDP per capita may have substantial pockets of poverty, and a region with low GDP per capita may have some areas of prosperity. The directives of the European Commission implicitly assume that the funding received by a region will be converted not only to greater prosperity on average, but will also reduce the existing disparities in the region (De Rynck and McAleavey, 2001). Resources awarded to a region whose average income level is low may simply result in additional well paid jobs for the narrow upper-middle class and, ultimately, in a greater inequality.

Inequality and growth are interlinked, but it is difficult to establish the direction of causality. Studies of the effect of growth on inequality traditionally refer to the hypoth-

[1]The EU15 comprised the following 15 countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom.

esis of Kuznets (1955). This hypothesis states that economic inequality increases over time while an area is developing until it reaches a certain level of per capita GDP and after that the inequality begins to decrease. At the same time, the level of inequality may affect, positively or negatively, the economic growth via distinct channels: accumulation of physical and human capital, redistributive public policies and political and social uncertainty (Weil, 2005). The regional (that is, sub-national) dimension enriches the debate on growth and income inequality, since one of the distinctive aspects of the regions is labour mobility. Perugini and Martino (2008) emphasize the role of regional dimension and they formulate a hypothesis about the association of the EU regional policies with inequality. They conclude that policies aimed at attracting mobile factors (capital and skilled labour) not only favour convergence, but also reduce the level of inequality in the poorer regions. On the other hand, Beckfield (2009) develops the idea that regional European policies succeed in reducing income disparities between EU member states, while simultaneously increasing economic inequality within European countries.

Monitoring income inequality as well as other indicators related to personal income distribution within European countries relies on comparable and internationally harmonized estimates at regional level for the member states. Harmonized household-level surveys are commonly used for aggregating personal income data by geographical areas of individual residences. Household surveys contain detailed income and related data, although sampling designs are usually not aimed for region-level estimates. For inferences about regional indicators from microdata of households there is always a trade-off between the (theoretically, historically, culturally and economically) appropriate level of territorial disaggregation to be adopted and the reduction of the sample size that affects the reliability of the estimates (Jesuit, 2008).

In the European Union, the Nomenclature of Territorial Units for Statistics (NUTS) is established by Eurostat and provides a single uniform breakdown of territorial units, generally defined in terms of the existing administrative units in the member states, for the production of regional statistics. In this hierarchical classification, each member state is subdivided at three levels: NUTS levels 1, 2 and 3 (Eurostat 2007). The member states may define further subdivisions of the NUTS3 units, such as local administrative units and municipalities[2].

The choice of the appropriate level for a socio-economic analysis of the European regions has been widely debated in the European Commission. The areas eligible for aid from the Structural Funds are the regions at NUTS2 level (Eurostat 2007), and therefore NUTS2 classification is the framework generally used by Member States for the application of their regional policies.

Comparable data on personal income distribution at the national and, *a fortiori*, at sub-national level are difficult to obtain. The Luxembourg Income Study (LIS) represents the most important project to assemble and *ex-post* harmonize microdata from income surveys conducted by national statistical agencies all over the world. The last wave of the LIS - wave VI, around 2004 - in addition to Unites States, Australia, Israel, Korea and other countries, includes 21 EU Member States.

The LIS database has been used only in a few studies of income inequality and poverty in the European Union at sub-national level. Jesuit et al. (2003) estimates regional poverty rates in five European countries and in Australia, Canada and United States, finding very high dispersion of poverty rates across regions, especially in Italy and the UK, and concludes that "the regional dimension is vitally important in measuring poverty" (p. 365). Stewart (2002) provides some regional indicators of well-being for the EU15 countries in the 1990's, emphasizing the uncertainty of the estimates even at NUTS1 level for some countries. Förster *et al.* (2005) concluded that the regional income inequality in the Eastern European countries in the 1990's was very high. Mahler (2002) provides measures of regional income inequality for a set of developed countries and relates income inequality to the turnout in national elections, complementing findings based exclusively on the analysis at the national level. Hoffmeister (2009) decomposes the dispersion of individual income in the late 1990's into its geographical components in 18 member states of EU25. With a few exceptions, the level of territorial disaggregation in these analyses is confined to NUTS1 units.

Focusing on the European Union, the pioneering European Community Household Panel (ECHP) survey, followed by its replacement, the annual European Statistics on Income and Living Conditions Survey (EU-SILC), is a principal source of data about socio-economic conditions of individuals and households in the EU countries and their

[2]For completeness of the classification, each country forms a NUTS0 unit.

regions. These national surveys are based on a standard questionnaire and provide *ex-ante* harmonized microdata on European countries. The ECHP was launched in 1994 in twelve member states (Belgium, Denmark, Germany, Greece, Spain, France, Italy, Ireland, Luxembourg, The Netherlands, Portugal, the United Kingdom). Austria joined the panel in 1995, Finland in 1996, and Sweden in 1997 with cross-sectional data derived from its National Survey of Living Conditions. Main characteristics of ECHP are cross-national comparability of the data and longitudinal dimension. Using data from ECHP, Rodriguez-Pose and Tselios (2009) mapped income per capita and inequality for 102 West European regions in 13 countries. The spatial unit of their analysis is that defined in ECHP and includes NUTS1 for 11 countries and NUTS2 for Portugal and the UK. To incorporate the recommendations of the UN "Canberra Manual" on household income definition and data collection and to improve data quality, ECHP was replaced in 2004 by the EU-SILC. The EU-SILC project is carried out under European Union legislation (council regulation No. 1177/2003) and it was formally launched in 2004 for the EU15. In 2006 EU-SILC covered the EU25 Member States as well as Norway and Iceland. The Survey is annual and it has a longitudinal and a cross-sectional component. Using data from EU-SILC, regional inequality has been studied within single countries (e.g. De Marco and Donatiello, 2008, for Italy), but there is no comprehensive study that covers all the EU countries.

This paper has two aims. First, based on the cross-sectional component of EU-SILC in 2006, we want to understand how EU-SILC can help final users to obtain information on personal income distribution at sub-national level. Second, we want to provide a comprehensive map on personal income distribution at country and regional level, as well as estimates of the Lorenz curves, Gini coefficients and poverty rates based on the equivalized household income in all the countries in which EU-SILC is conducted. We also estimate inequality and poverty at regional level NUTS1 and NUTS2, when they are recorded in the database. Since European directives on EU-SILC impose only national representativeness, sample sizes might not assure reliable estimates at sub-national level. Therefore, we apply a simple method of small-area estimation to improve the regional estimates. To indicate the uncertainty of the estimates, we estimate their standard errors by bootstrap method.

The paper is organized as follow. Section 2 gives details of the data and the definitions of the inequality measures we estimate. Section 3 provides a map of inequality in the European countries along with a brief description of the bootstrap procedure we implemented for assessing the uncertainty of non-linear statistics. Section 4 reports inequality and poverty measures in the European regions and an outline of the method for small-area estimation for regions with small samples in the data. Section 5 discusses the results and draws conclusions.

2 Data and definitions

Table 1 lists the countries, their sample sizes, the level of disaggregation (NUTS1 or NUTS2) recorded in database[3], and estimated population size in 2006. Territorial areas are recorded only for some countries. Some countries are single units at both NUTS1 and NUTS2 levels.

In all analyses, we use the equivalized household income, and after its definition below we refer to it for brevity as "income". Household is defined as a person living alone or a group of persons living together in the same dwelling, sharing expenditure and having the joint provision of the essentials for living. Persons living in collective households and in institutions are generally excluded from the target population.

The income (without equivalization) of a household is defined as the sum of the personal income components of all household members and income components associated with the household as a whole. Personal income includes cash and non-cash employee income, income from self-employment, financial transfers received from outside the household (including social benefits). Income components of the household that are not associated with any single member include property income (rental and interest/dividend from capital, imputed rent) and housing allowances. Total income is net of income tax and social contributions. The following payments are subtracted from this total: interest paid on mortgage(s), taxes related to wealth, inheritance, purchase of residential property, and the like, and certain financial transfers from household members to individuals or institutions outside the household. Until 2006, interest on mortgage was not recorded and the methods used by the member countries were not

[3]Cross-sectional UDB SILC 2006 Rev.1, March 2009.

Table 1: Information about the countries in the EU-SILC (2006). The number of regions at the level of disaggregation is given in parentheses.

	Country	Sample size	Level of disaggregation	Population size (in mil.)
At	Austria	6028	NUTS1 (3)	8.3
Be	Belgium	5860	NUTS1 (3)	10.5
Cy	Cyprus	3621	NUTS1 (1)	0.8
Cz	Czech Republic	7483	NUTS2 (8)	10.3
Dk	Denmark	5711	NUTS1 (1)	5.4
Ee	Estonia	5631	NUTS1 (1)	1.3
Fi	Finland	10 868	NUTS2[1] (4)	5.3
Fr	France	10 036	NUTS2[2] (22)	63.2
De	Germany	13 799	NUTS1[3] (6)	82.4
Gr	Greece	5700	NUTS1 (4)	11.1
Hu	Hungary	7722	NUTS1 (3)	10.1
Is	Iceland	2845	NUTS1 (1)	0.3
Ei	Ireland	5836	NUTS1 (1)	4.3
It	Italy	21 499	NUTS2[4] (21)	58.9
Lv	Latvia	4315	NUTS1 (1)	2.3
Lt	Lithuania	4660	NUTS1 (1)	3.4
Lu	Luxembourg	3836	NUTS1 (1)	0.5
Nl	The Netherlands	8986	NUTS0 (1)	16.3
No	Norway	5765	NUTS1 (1)	4.7
Pl	Poland	14 914	NUTS1 (6)	38.1
Pt	Portugal	4367	NUTS0 (1)	10.6
Sk	Slovakia	5105	NUTS1 (1)	5.4
Si	Slovenia	9478	NUTS0 (1)	2.0
Es	Spain	12 205	NUTS2 (19)	44.1
Se	Sweden	6803	NUTS1 (1)	9.1
UK	United Kingdom	9902	NUTS0 (1)	60.6

Notes:

[1] Region Åland not represented in the data.

[2] Overseas departments are not represented in the survey.

[3] Recoded to fewer regions than the orginal NUTS1 classification to 16 regions.

[4] The disaggregation was obtained from the national release of the data.

fully harmonized. Therefore we do not include the imputed rents and the interest on mortgage in the total household income.

The equivalized household income is obtained by dividing the total household income by its modified OECD equivalence scale. According to this scale, the first adult member of the household counts as 1.0, and all other adult members as 0.5. Each child (below 14 years of age) is accorded the weight 0.3. For example, a household with four adults and two children has equivalized size $1 + 3 \times 0.5 + 2 \times 0.3 = 3.1$, so that if their total income is 62 000 Euro, then the equivalized household income is 20 000 Euro[4].

In the following, we briefly discuss the measures that we apply for the estimation both at the country and region level, namely Lorenz curve, Gini coefficient and poverty rate. The Lorenz curve is defined as the relative partial integral of the expectation of the income. That is, let $f(x)$ be the probability density of the income, F its cumulative distribution function,

$$F(y) \;=\; \int_0^y f(x)\,\mathrm{d}x\,,$$

and $\mathrm{E}(X) = \int_0^{+\infty} x f(x)\,\mathrm{d}x$ its expectation. Then the Lorenz curve at quantile $p \in (0,1)$ is defined as

$$L(p) \;=\; \frac{1}{\mathrm{E}(X)} \int_0^{F^{-1}(p)} x f(x)\,\mathrm{d}x\,,$$

or equivalently as

$$L(p) \;=\; \frac{1}{\mathrm{E}(X)} \int_0^{p} F^{-1}(r)\,\mathrm{d}r\,.$$

We refer to the argument p as the proportion (and to $100p$ as the percentage) of the households, and to $L(p)$ as the accumulated income. Every Lorenz curve lies between the zero and identity lines ($y = 0$ and $y = p$, respectively). They correspond to the extreme settings of each household having the same income (perfect equality) and a single household having all the income (extreme inequality).

The Gini coefficient is defined as the fraction of the area between the equivalence (identity) line $g(x) = x$ and the Lorenz curve:

$$G \;=\; 1 - 2\int_0^1 L(x)\,\mathrm{d}x\,.$$

The Gini coefficient (for a country or a region) is estimated by numerical integration of the estimated Lorenz curve. Numerical quadrature, which is used for this purpose,

[4]We decided not adjust income figures for Purchasing Power Parity (PPP) since PPPs at regional level are not available for European countries.

entails an approximation error. Further error is incurred because the Lorenz curve is estimated from the survey data.

The poverty rate for a country is defined as the percentage of households whose income falls short of 60% of the national median income, according to the Eurostat definition. We estimate this quantity also for regions, but use the national median as the standard.

Each household in the data is associated with a sampling weight. All the estimates take these weights into account. For example, the Lorenz curve at a point p is estimated by the following steps:

1. sort the households in the survey in the increasing order of their income;
2. form the cumulative totals of the weights of the sorted households;
3. identify the household for which the cumulative total of weights is equal to $100p\%$ of the overall total of the weights;
4. the Lorenz curve at p is defined as the corresponding cumulative total of (sorted) income, divided by the weighted total of income.

As $100p\%$ of the total of the weights is not matched with the cumulative total for any household, the first household for which the cumulative total exceeds this percentage contributes to the value of the Lorenz curve only with half of its income.

The Lorenz curve is estimated on a dense grid of values of p. For countries, we use the percentile grid, with $p = 0.01, 0.02, \ldots, 0.99$. For regions, we use the grid with 30 equidistant points, to avoid problems caused by relatively small sample sizes for some regions. The values of the Lorenz curve at 0 and 1 are equal to 0 and 1 respectively. Households with negative income are included in the analysis.

3 Measuring poverty and inequality in the European Union

In this section we provide details of inequality and poverty measures for all the European countries for which EU-SILC is conducted. Income refers to equivalized household income as defined in Section 2. We provide estimates of median income, Gini coefficient, the value of the Lorenz curve at $p = 0.50$ and the poverty rate (with 60% of the median income as the reference). We deal with sampling design effects by applying the cross-sectional sampling weights associated to each household in the survey.

3.1 Estimating uncertainty of inequality and poverty measures

Standard errors of the estimated median incomes, Gini coefficients, Lorenz curves and poverty rates are assessed by a *bootstrap* procedure. Bootstrap is highly recommended for estimating sampling variation of complex non-linear statistics such as the Gini coefficient, because no analytical expressions are available (see, among others, Mills and Zandvakili, 1997; Biewen, 2002).

The bootstrap procedure we implemented draws a large number of independent samples with replacement from the original sample, each with the same sample size as the original sample and it accounts for the cross-sectional sampling weights . The draw of a bootstrap sample with sampling weights consists of the following steps:

1. Each household is represented by a segment of length equal to its sampling weight. The segments are joined to form a single segment. Let its length be Λ.

2. Let n' be the effective size of the original sample. It is calculated from the sampling weights w_i as

$$n' = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2} .$$

   A random sample Ξ of size n' from the uniform distribution on $(0, \Lambda)$ is drawn. A household is selected into the bootstrap sample h times if h of the n' elements of Ξ fall into its segment.

A schematic example is given in Figure 1. We simplify this scheme by drawing separate bootstrap samples for each country (or region) in the data, assuming that the sampling design is stratified by the countries (or regions). The effective sample size (Potthoff, Manton and Woodbury, 1993) is an approximation to the size of a simple random sample that would have the same amount of information as the original sample. The effective sample size is smaller or equal to the actual sample size, and equality occurs only when the sampling weights are constant.

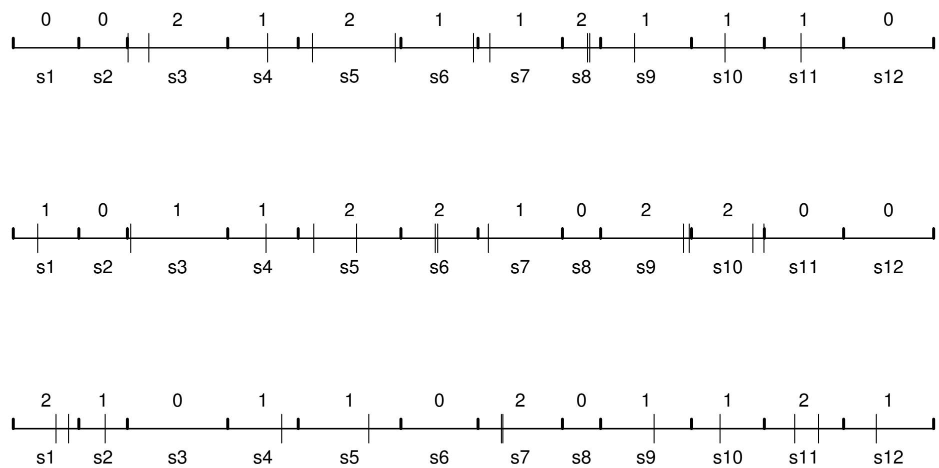

Figure 1: Schematic example of three replications of the bootstrap for 12 observations $(s1, s2, \ldots, s12)$ with weights. The short ticks mark the cumulative weights of the observations, and the long thin ticks the random draws from the uniform distribution on $(0, \Lambda = \sum_i w_i)$. The numbers of selections of each observation are given above the segments.

As previously pointed out, bootstrap is particularly recommended for estimating sampling variation of non-linear statistics for national and regional (small-area) quantities (see Section 4.1).

In our bootstrap implementation we apply the estimator of the Lorenz curve and the Gini coefficient to each of these bootstrap samples, with the sampling weights set to unity. This results in a bootstrap estimate for each sample. The standard error of the estimator of the Gini coefficient is estimated by the standard deviation of the bootstrap estimates. The bias of the estimator is estimated by the difference between the mean of the bootstrap estimates and the original estimate.

3.2 Inequality and poverty in the European Countries

Table 2 reports the estimates along with their bootstrap standard errors of median income, Gini coefficient, the value of the Lorenz curve at $p = 0.50$ and the poverty rate (with 60% of the median income as the reference) for all the European countries participating the EU-SILC. For easier comprehension, the values of the Gini coefficient and $L(0.5)$ are multiplied by 100. The countries are divided into three groups according to their median income with cutpoints of 10 000 and 18 000 Euro: low income, medium income and high median income. All the estimates are averaged over 100 bootstrap samples, so as to avoid the bias of the direct estimator. The standard errors are also estimated by the weighted bootstrap procedure (see Section 3.1).

The estimates are plotted in Figure 2 using distinct symbols for the three groups of countries. The strong negative association of the Gini coefficient and $L(0.5)$ is a direct consequence of the definition of the Gini coefficient in terms of the Lorenz curve. High median income is generally associated with low Gini coefficient and high $L(0.5)$. The countries with low median income have a wide range of values of the Gini coefficient (26–40%), whereas the range for the other countries is much narrower (25–34%).

Inequality differs substantially across European countries, ranging from 25% (Sweden) to more than 40% (Portugal). Although our analysis cannot be considered a formal assessment, our data do not provide any support for a systematic relationship between inequality and income. It seems that the Kuznets curve that works for a cross-section of countries at a particular point in time (Anand and Kanbur, 1993; Li, Squire and Zou, 1998), does not apply to our subset of European countries.

Countries that belong to the low median income group have, in general, high poverty rates (18% or more). Latvia stands with the lowest median income, but the highest estimated poverty rate. There are some notable exceptions to this rule: Slovakia, with a slightly higher median income than Latvia, and Czech Republic, with about twice as high median income as Latvia, have very small poverty rates (estimates 8.33% and 9.85%).

Most of the countries in the medium income group have very low poverty rates (9 – 14%). The Netherlands has the second lowest estimated poverty rate (9.05%). Some exceptions are Cyprus, Italy and Spain with poverty rate ranging from 18% to 21%.

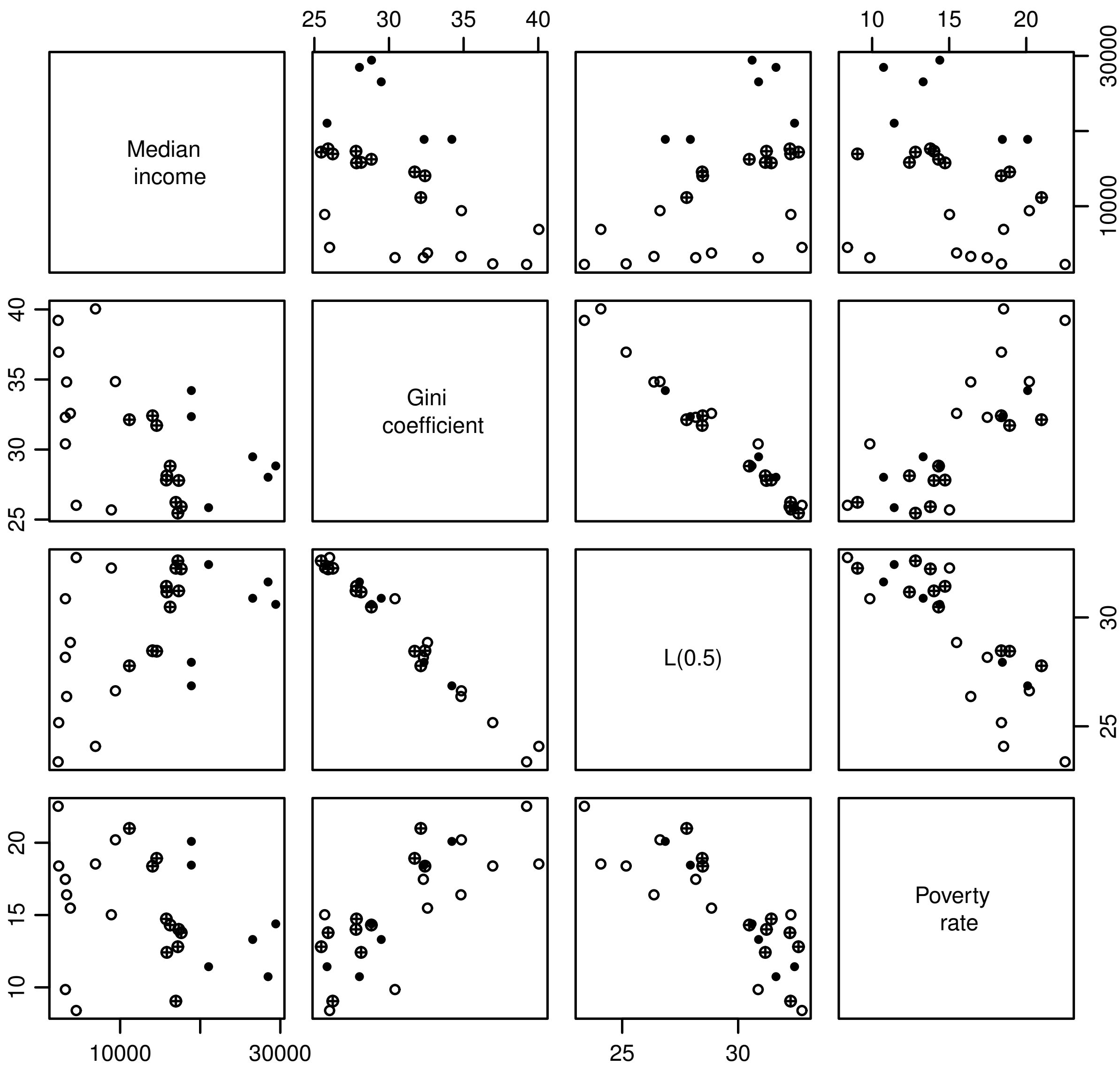

Figure 2: Matrix plot of the estimated national indicators of poverty and inequality. The symbols denote the three groups of countries indicated in Table 2: ○ — low median income; ⊕ — medium median income; • — high median income.

Among the countries with high median income, the UK stands out with a poverty rate greater than 20%.

The highest values of the Gini coefficient occur for the countries with low median income (Portugal, Latvia and Lithuania), but Czech Republic and Slovenia, also in this income group, have low Gini coefficients. Italy and Spain are the most unequal among the countries belonging to the medium group (31.72% and 32.13% respectively), characterized also by high poverty rates. In the high median income group, the UK and Ireland have the highest degree of inequality (Gini coefficient equal to 34.21% and 32.35% respectively) and with the highest levels of poverty rates (18.45% and 20.09%).

Given the skewness of the income distribution within the countries, the standard errors of the median income are strongly associated with the level of median income.

4 Analysis of regions

The countries in which EU-SILC was conducted in 2006 differ a lot in terms of inequality and poverty. They also differ in size, and therefore in their numbers of NUT2 units, between a single unit for Cyprus, Iceland, Latvia and Luxembourg to 39 units in Germany. Any comparisons of the differences among the regions in one country with the differences in another has to be interpreted with caution. Variation of inequality and poverty across regions within countries is often wider than the variation across countries and the level of disparity is likely to be greater in countries with more regions. In any case, EU-SILC micro data that would enable us to identify regions and compare them are available only for a few countries. No location is indicated in the data for the UK and Poland, and it is collapsed to only six categories in Germany.

In this section, we present the estimated measures of regional inequality and poverty along with their relative standard errors for all the countries that have an appreciable number of NUTS2 regions identified in the EU-SILC database, namely France, Italy, Spain and Czech Republic[5].

[5] Estimates of regional inequality and poverty at NUTS1 level for the other countries are available on request.

4.1 Estimating inequality and poverty in regions with small sample size

A drawback in estimating inequality measures and poverty rates at a more detailed geographical level is the lack in precision due to small sample sizes of the territorial units. For countries with many regions in the data, France, Italy and Spain, the sample sizes for some of the regions are too small for reliable estimation of their quantities. We deal with this problem, by applying a simple method for small-area estimation that improves the estimates by exploiting the similarity of the regional quantities. The method can be motivated and formalized as follows.

Suppose a country is divided into D regions and the (unknown) quantities associated with them are θ_d, $d = 1, \dots, D$. Their national counterpart is denoted by θ. Suppose $\hat{\theta}_d$ is an unbiased estimator for region d, based solely on the income of the sampled households in the region. Such an estimator is called *direct*. Let its sampling variance be v_d. Let $\hat{\theta}$ be an unbiased estimator of the national parameter θ, and v its sampling variance.

For a region with a small sample size we have two candidate estimators. The estimator based on the data for the region is unbiased, but it has a large sampling variance. The national estimator is biased for the region, but has a small variance because it is based on a large sample. We might select the estimator that has, or is believed to have, smaller mean squared error (MSE), but it is more effective to combine the two estimators. This can be interpreted as *shrinking* the unbiased director estimator toward the stable (small-variance) national estimator. The extent of shrinkage is set by aiming to minimize the MSE of the combination.

Formally, a small-area estimator of θ_d is defined as a convex combination of $\hat{\theta}_d$ and $\hat{\theta}$,

$$\tilde{\theta}_d \;=\; (1 - b_d)\,\hat{\theta}_d + b_d\hat{\theta}\,,$$

where b_d is a constant specific for the region d as well as the parameter θ_d. Its value is set so as to minimise the MSE of the combination. This MSE is

$$(1 - b_d)^2 v_d + 2b_d(1 - b_d)c_d + b_d\left(v + B_d^2\right)\,,$$

so the optimal coefficient b_d is

$$b_d^* \;=\; \frac{v_d - c_d}{v_d + v - 2c_d + B_d^2}\,, \tag{1}$$

where $B_d = \mathrm{E}(\hat{\theta}_d) - \theta$ is the bias of $\hat{\theta}_d$ in estimating θ_d and $c_d = \mathrm{cov}(\hat{\theta}_d\,, \hat{\theta})$.

The squared deviation B_d^2 is not known and would be estimated with only modest precision, so we substitute for B_d^2 its average over the regions:

$$\sigma_{\mathrm{B}}^2 \;=\; \frac{1}{D}\sum_{d=1}^{D} (\theta_d - \theta)^2\;.$$

This (region-level) variance itself has to be estimated, but that can be done with greater precision than most values of B_d^2. We apply moment matching using the statistic

$$S_{\mathrm{B}} \;=\; \sum_{d=1}^{D} \left(\hat{\theta}_d - \hat{\theta}\right)^2\;.$$

Its expectation is

$$\begin{aligned}
\mathrm{E}\left(S_{\mathrm{B}}\right) \;&=\; \sum_{d=1}^{D} (\theta_d - \theta)^2 + \sum_{d=1}^{D} (v_d - 2c_d + v) \\
&=\; D\sigma_{\mathrm{B}}^2 + Dv + \sum_{d=1}^{D} (v_d - 2c_d)\;.
\end{aligned}$$

Hence the moment matching estimator

$$\hat{\sigma}_{\mathrm{B}}^2 \;=\; \frac{1}{D}\left\{S_{\mathrm{B}} - \sum_{d=1}^{D}(v_d - 2c_d)\right\} - v\,. \tag{2}$$

It can be interpreted as a bias correction of S_{B}/D. When there are many regions D and the sample of none of them forms a large part of the overall sample, $v \ll v_d$ and $c_d \ll v_d$ for all the regions. Then, c_d and v can be dropped from the expression in (2), if simplicity is preferred to a small loss of efficiency. When $\hat{\theta}$ is a linear combination of $\hat{\theta}_d$, $\hat{\theta} = (w_1\hat{\theta}_1 + \cdots + w_D\hat{\theta}_D)/w_+$, the covariances are $c_d = w_d/w_+ v_d$ and $v = (w_1^2 v_1^2 + \cdots + w_D^2 v_D^2)/w_+^2$. These expressions are good approximations even when $\hat{\theta}$ is not a linear function of the $\hat{\theta}_d$. Thus, the ideal coefficient b_d^* is estimated by

$$\hat{d}^* \;=\; \frac{\hat{v}_d\left(1 - \dfrac{w_d}{w_+}\right)}{\hat{v}_d\left(1 - 2\dfrac{w_d}{w_+}\right) + v + \hat{\sigma}_{\mathrm{B}}^2}\,,$$

and the original target θ_d by

$$\tilde{\theta}_d \;=\; \left(1-\hat{b}^*_d\right)\hat{\theta}_d + \hat{b}^*_d\hat{\theta}_d\,.$$

This estimator is less efficient than $\tilde{\theta}_d(b^*_d)$, if we could obtain the latter, because we estimate the variances v_d, v and σ^2_{B} and because we substitute σ^2_{B} for B^2_d. The composite estimator $\tilde{\theta}_d$ is usually more efficient than $\hat{\theta}_d$ for most of the regions. It is less efficient for regions with absolute deviations $|\,B_d\,|$ much greater than the standard deviation σ_{B}. Of course, these regions cannot be identified from the data, especially when they have only moderate or small subsample sizes.

The MSE of $\tilde{\theta}_d$ is estimated by the attained minimum, which is equal to

$$v_d - \frac{(v_d-c_d)^2}{v_d-2c_d+v+\sigma^2_{\mathrm{B}}}\,.$$

An improvement of this estimator is proposed in Longford (2005).

The small-area estimator $\tilde{\theta}_d$ differs from $\hat{\theta}_d$ only slightly for most regions because they have substantial sample sizes. However, for a few regions with small sample sizes it is very useful[6].

4.2 Inequality and poverty in European regions

Estimated inequality measures and poverty rates for the NUTS2 regions of France, Italy, Spain and Czech Republic are displayed in the respective Tables 3 – 6, along with their estimated standard errors. The regions of the former three countries are split into geographical groups; the groups have no official status, although for Italy they coincide with a commonly used division.

Figures 3 – 6 display the sets of Lorenz curves for the regions of the four countries, with the insets in the right-hand panels listing and identifying the regions. The Gini coefficients for the regions of a country can be summarized by their averages and estimated variances. The latter are the estimates $\hat{\sigma}^2_{\mathrm{B}}$ given by (2), estimating without bias

[6]Note that we do not address the representativeness issue. In principle, a re-weighting scheme could be applied, post-stratifying the regional sub-samples to obtain the marginal distributions of households by several socio-economic groups, provided that population and, for instance, labour force statistics are available at regional level. In this way, we could obtain more accurate regional direct estimates of the indicators of interest. This we leave for future research, since here we focus on dealing with regions that have small subsamples.

the variance of the regions' population Gini coefficients. The estimates of the standard deviations $\hat{\sigma}_{\mathrm{B}}$ for France, Italy and Spain are 2.43, 2.24 and 2.96%, respectively. For all three countries, these are estimates for NUT2 regions, but their comparison is problematic because of various anomalies in how the regions are defined and data collected. In particular, two regions of Spain, Ceuta and Melilla, are relatively small cities at the Moroccan (African) coast. They have large Gini coefficients. No data was collected from the overseas departments of France, and Corsica is represented by only 24 subjects, too few for estimating the Lorenz curve with any appreciable precision. The sample sizes for the other regions are in the range 149–1690, largely reflecting the population sizes of the regions. In Spain, the two African cities have sample sizes 110 (Melilla) and 138 (Ceuta); they are substantially oversampled. The population sizes of the regions of Czech Republic are in a very narrow range (1.1–1.5 million) and are on average much smaller than in the other countries.

Different patterns of regional inequality and poverty within the four selected countries are detected. In France and Czech Republic, there are similar findings in that income inequality is the highest in the areas with capital cities (Île de France and Praha where Paris and Prague belong to) and well above the national figures (France Gini=28.11; Île de France Gini=30.22; Czech Republic Gini=26.06; Praha Gini=31.63). This evidence comes along with the fact that Île de France and Praha are the richest regions in the country as well. Disparity is present also in Italy, but to a lesser extent (Italy Gini=31.72; Lazio Gini=32.92); Lazio is not the richest region in Italy. Inequality in Italy is highest in the poorest regions of the South (Campania and Sicilia). In Spain, in contrast, income inequality in Madrid, its capital, is lower than the national figure (Madrid Gini=30.80; Spain Gini=32.13). The poverty rates in all the regions with capital cities are much smaller than the corresponding country figures (France Poverty rate=12.52; Île de France Poverty rate=11.29; Czech Republic Poverty rate=8.33; Praha Poverty rate=3.77; Italy Poverty rate=18.92; Lazio Poverty rate=17.90). This finding is not surprising since the poverty lines are fixed as 60% of the national median income and the capital regions are among the richest within the country.

The Lorenz curves for the eight regions of Czech Republic are dispersed much less than the curves for the other three countries. Such a comparison can be made only based on the left-hand panels of Figures 3 – 6, because the axes in the insets (the right-hand panels) do not have identical scales. The curve for region 1 (Praha, the capital) has the smallest value throughout the range (0,100).

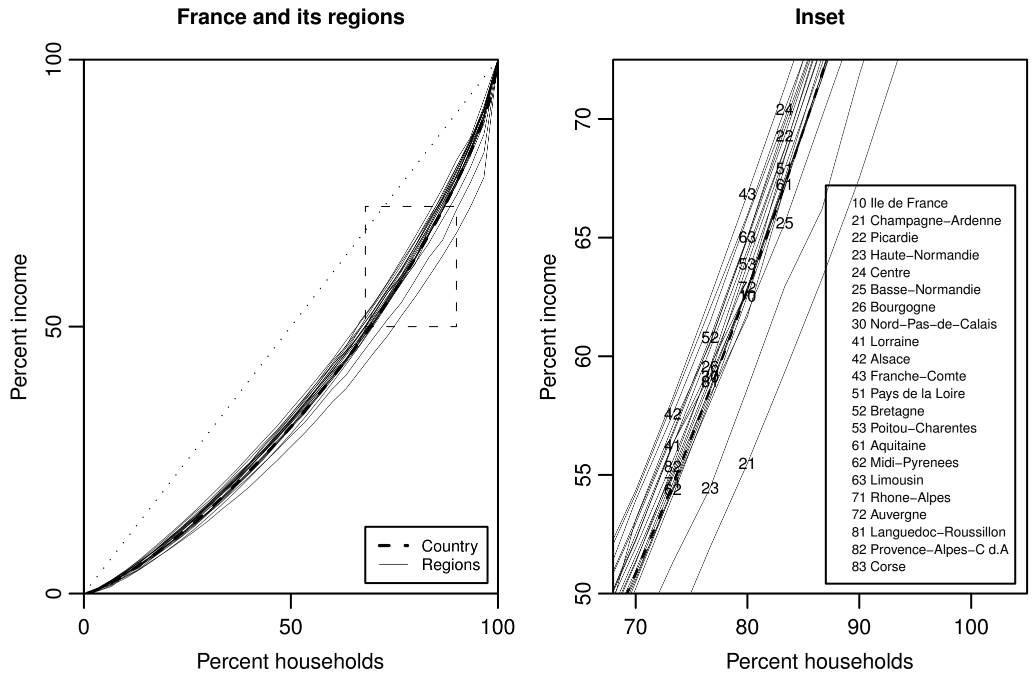

Figure 3: The Lorenz curves for the NUTS2 regions of France. The right-hand panel is a zoom-in on the rectangle marked in the left-hand panel.

5 Conclusions

European Statistics on Income and Living Conditions (EU-SILC) is a reference source for comparative statistics on income, poverty and social exclusion. The survey has been first carried out for EU15 in 2004 and, from 2005, for all the EU member states. The survey design is aimed to obtain representative samples at both EU and country levels, as well as for several subgroups, such sex, household size, household type and socio-economic group. To promote uniformity of regional policy evaluations, Eurostat recommends that income related analysis be based on data from EU-SILC.

A major drawback of the survey is that its sample sizes and designs are not sufficient for reliable direct estimation for all regions within countries. Besides concerns about confidentiality preclude recording of some geographical information altogether,

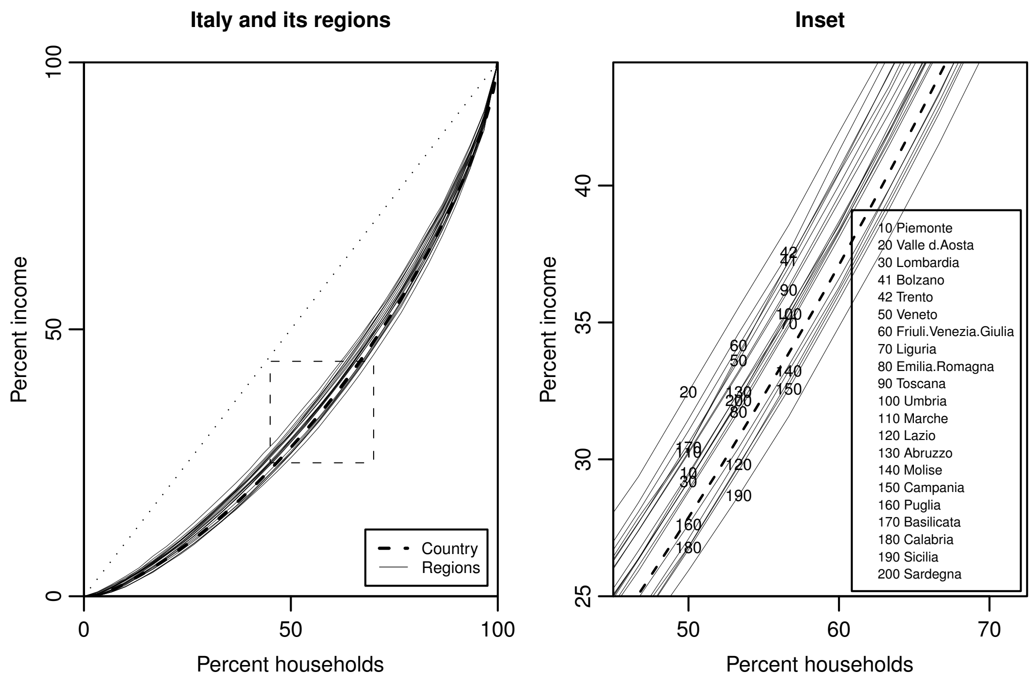

Figure 4: The Lorenz curves for the NUTS2 regions of Italy. The right-hand panel is a zoom-in on the rectangle marked in the left-hand panel.

the sample size, especially in those where the level of disaggregation is NUTS2, is too small to generate accurate direct estimates for detailed level of disaggregation, with the consequence that standard errors associated to the estimates could be excessively large. However, policy makers require reliable estimators of economic indicators, such as poverty or income inequality, at sub-national level in order to monitor a propose policy program that speed up development of the more disadvantaged areas in Europe. We applied small-area estimation (SAE) techniques to obtain more accurate estimate of summaries related to NUTS2 units. The basic idea is to combine the struggles of the regional direct estimates, that are, in principle, unbiased but have high standard errors, and the national level estimate, which is biased for every region but it is very stable. We set the weights/coefficients in the convex linear combination of the regional direct estimates and the national estimates with the intent to minimize the minimum mean square error. Uncertainty of the inequality and poverty figures is accessed by a bootstrap procedure that takes into account also sampling weights.

We analyzed income figures (median income), poverty (poverty rates) and inequality

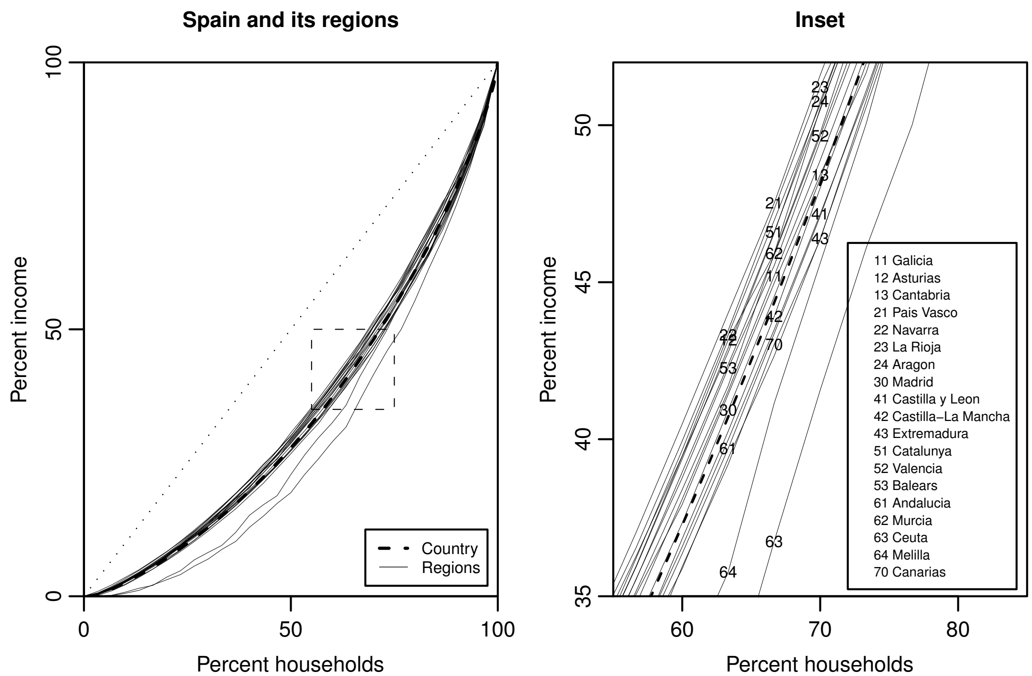

Figure 5: The Lorenz curves for the NUTS2 regions of Spain. The right-hand panel is a zoom-in on the rectangle marked in the left-hand panel.

(both Gini coefficient and Lorenz curve) for all the European countries for which EU-SILC data are available, along with their bootstrapped standard errors. The largest value of the Gini coefficient is in Portugal, Latvia and Lithuania, countries that are characterized by by low median income. The lowest value of median income is found in Latvia, while the highest value is in Luxembourg. Latvia stands out also for its highest estimated poverty rate, Czech Republic in contrast is characterized by the smallest estimated poverty rate, even though its median income is quite low.

At sub-regional level we used small area estimation to measure median income, poverty and inequality in France, Italy, Spain and Czech Republic, countries characterized by small regional sub-samples. Out method was quite effective for regions with small sample sizes and allowed us to provide an exhaustive map of income distribution, poverty and inequality also within European countries. Our findings suggest that regional factors play an important role in shaping inequality in Europe. The bulk of inequality within EU countries is accounted for by intra-countries inequality rather than between-countries, giving evidence that allocation resources at regional level may

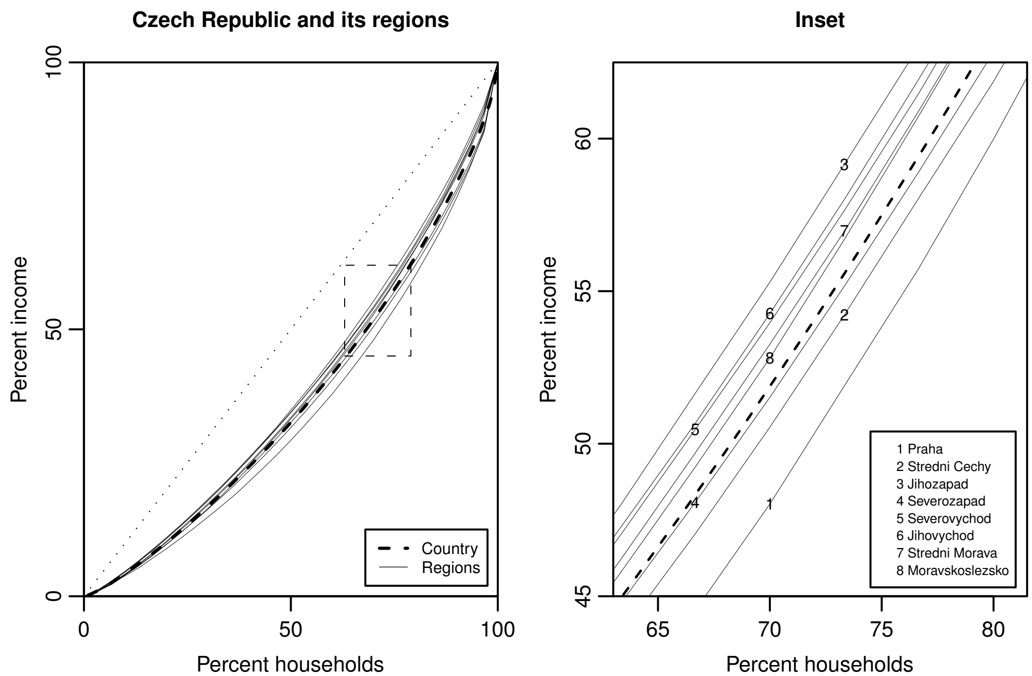

Figure 6: The Lorenz curves for the NUTS2 regions of Czech Republic. The right-hand panel is a zoom-in on the rectangle marked in the left-hand panel.

contribute to reduce inequality and poverty towards a more harmonious convergence between European regions.

Interestingly, Italy and France have regions belonging to "extreme" categories in terms of poverty and inequality; the other countries for which regional analysis was possible (Spain and Czech Republic) are more homogenous. Unfortunately, for some other countries (such as Portugal and the UK) the public file does not release any information about regions.

Finally, we have proposed a new approach for estimating poverty and inequality at regional level along with their associated sampling variation. The proposed procedure is particularly recommended when we estimate non linear statistics (such as Gini coefficient), when we deal with areas with small sample sizes and when we want to account for sampling weights.

Table 2: Estimates and relative standard errors of the national indicators of income level and inequality in the year 2006.

	Median income		*Gini coefficient*		$L(0.5)$		*Poverty rate*	
	Estimate	St. error	Estimate	St. error	Estimate	St. error	Estimate	St. error
Low median income								
Cz	4517	(34)	26.06	(0.44)	32.71	(0.23)	8.33	(0.32)
Ee	3318	(48)	34.82	(0.58)	26.37	(0.34)	16.40	(0.85)
Gr	9410	(105)	34.85	(0.54)	26.63	(0.30)	20.20	(0.62)
Hu	3778	(35)	32.58	(0.58)	28.85	(0.30)	15.48	(0.27)
Lv	2256	(39)	39.22	(0.59)	23.37	(0.33)	22.52	(0.96)
Lt	2330	(33)	36.95	(0.51)	25.17	(0.32)	18.39	(0.95)
Pl	3152	(20)	32.30	(0.27)	27.17	(0.16)	17.47	(0.24)
Pt	6919	(87)	40.04	(0.85)	24.08	(0.45)	18.53	(0.62)
Sk	3164	(25)	30.40	(1.22)	30.86	(0.57)	9.85	(0.35)
Si	8899	(52)	25.69	(0.28)	32.27	(0.18)	15.02	(0.42)
Medium median income								
At	17 630	(133)	25.92	(0.31)	32.23	(0.19)	13.78	(0.51)
Be	16 242	(118)	28.82	(0.62)	30.48	(0.31)	14.31	(0.50)
Cy	14 046	(156)	32.41	(0.91)	28.47	(0.44)	18.38	(0.74)
Fi	17 316	(171)	27.79	(0.70)	31.22	(0.34)	14.01	(0.54)
Fr	15 828	(102)	28.11	(0.30)	31.19	(0.16)	12.52	(0.33)
De	15 771	(58)	27.82	(0.47)	31.43	(0.25)	14.73	(0.29)
It	14 560	(69)	31.72	(0.26)	28.45	(0.15)	18.92	(0.25)
Nl	16 956	(126)	26.23	(0.51)	32.26	(0.28)	9.05	(0.34)
Es	11 155	(118)	32.13	(0.37)	27.78	(0.21)	20.99	(0.52)
Se	17 200	(97)	25.46	(0.39)	32.60	(0.22)	12.81	(0.34)
High median income								
Dk	21 044	(163)	25.85	(0.54)	32.43	(0.30)	11.43	(0.44)
Is	28 469	(262)	28.02	(1.08)	31.63	(0.54)	10.74	(0.50)
Ei	18 887	(317)	32.35	(1.03)	27.94	(0.48)	18.45	(1.19)
Lu	29 427	(569)	28.83	(0.68)	30.60	(0.38)	14.39	(1.22)
No	26 547	(279)	29.48	(1.58)	30.88	(0.79)	13.31	(0.56)
UK	18 893	(143)	34.21	(0.41)	26.86	(0.21)	20.09	(0.41)

Note: see Table 1 for the acronyms used for the countries.

Table 3: Estimates of the French regional indicators of income level and inequality. The Gini coefficient for Corse (Corsica) is not estimated because of small sample size, 24.

Region	*Population size* (in '000)	*Median income*		*Gini coefficient*		*L*(0.5)		*Poverty rate*	
		Estimate	St.err.	Estimate	St.err.	Estimate	St.err.	Estimate	St.err.
North									
Île de France	11 574.4	19 283	(252)	30.22	(0.78)	28.91	(0.54)	11.29	(0.80)
Champagne-Ardenne	1337.4	14 607	(466)	26.91	(2.20)	31.06	(2.12)	13.38	(2.34)
Picardie	1896.2	14 644	(314)	24.29	(1.45)	35.04	(0.69)	11.01	(1.82)
Haute-Normandie	1812.0	15 816	(568)	27.89	(2.26)	29.66	(1.70)	11.36	(2.26)
Centre	2524.5	16 094	(300)	23.51	(1.16)	35.63	(0.64)	7.11	(1.14)
Basse-Normandie	1458.4	14 970	(413)	26.23	(1.96)	32.89	(1.26)	12.46	(1.86)
Bourgogne	1629.4	15 798	(423)	24.09	(1.16)	33.67	(0.86)	12.20	(1.89)
Nord – Pas-de-Calais	4020.1	14 621	(309)	27.14	(1.32)	32.33	(0.69)	12.96	(1.37)
East									
Lorraine	2336.1	14 436	(299)	25.06	(0.99)	33.65	(0.64)	13.11	(1.82)
Alsace	1820.7	16 544	(316)	23.38	(1.31)	35.52	(0.75)	9.62	(1.92)
Franche-Comté	1152.6	15 553	(785)	23.25	(0.95)	34.68	(0.84)	9.85	(1.64)
West									
Pays de la Loire	3465.4	14 823	(333)	26.37	(1.03)	33.30	(0.62)	14.31	(1.71)
Bretagne	3106.5	16 176	(276)	23.73	(1.06)	34.25	(0.59)	8.98	(1.17)
Poitou-Charentes	1729.1	15 663	(464)	24.60	(1.37)	33.34	(0.77)	12.77	(2.05)
Aquitaine	3133.1	15 196	(347)	28.34	(1.08)	29.69	(0.73)	19.07	(2.10)
Midi-Pyrénées	2791.4	15 535	(305)	27.73	(1.14)	30.81	(0.81)	13.24	(2.14)
Limousin	732.0	15 542	(331)	23.71	(1.36)	33.86	(1.12)	16.38	(2.47)
South									
Rhône-Alpes	6047.4	16 521	(251)	27.69	(0.97)	31.11	(0.58)	9.58	(0.97)
Auvergne	1337.5	14 124	(445)	26.34	(1.68)	34.25	(0.99)	9.79	(2.39)
Languedoc-Roussillon	2549.6	14 620	(418)	26.22	(1.37)	33.41	(0.83)	15.27	(1.81)
Provence-Alpes-C.d.A	4835.1	14 894	(288)	26.90	(0.75)	31.54	(0.52)	16.14	(1.72)
Corse	296.3	13 283	(1146)	—	(—)	32.45	(1.92)	17.36	(5.25)

Table 4: Estimates of the Italian regional indicators of income level and inequality.

Region	*Population size* (in '000)	*Median income*		*Gini coefficient*		$L(0.5)$		*Poverty rate*	
		Estimate	St.err.	Estimate	St.err.	Estimate	St.err.	Estimate	St.err.
North									
Piemonte	4347.3	15 372	(250)	30.32	(0.70)	29.49	(0.56)	14.09	(1.28)
Valle d'Aosta	124.4	15 982	(236)	26.24	(1.29)	35.68	(0.78)	9.39	(1.73)
Lombardia	9510.3	16 523	(203)	30.30	(0.70)	29.36	(0.46)	11.59	(0.89)
Bolzano	485.2	16 164	(697)	29.74	(1.31)	28.46	(0.64)	16.54	(2.35)
Trento	504.8	17 262	(340)	26.75	(1.12)	32.74	(0.69)	8.10	(1.66)
Veneto	4755.9	15 414	(284)	29.16	(0.93)	30.58	(0.40)	13.78	(0.70)
Friuli-Venezia Giulia	1210.4	15 936	(252)	28.21	(1.00)	32.79	(0.48)	13.83	(1.48)
Liguria	1609.0	14 866	(351)	29.63	(0.72)	29.65	(0.43)	14.38	(1.76)
Centre									
Emilia-Romagna	4205.4	17 039	(261)	30.41	(0.82)	30.25	(0.46)	11.23	(0.97)
Toscana	3629.0	16 402	(205)	30.07	(0.88)	29.14	(0.45)	10.58	(1.11)
Umbria	870.4	14 184	(474)	30.46	(0.87)	28.35	(0.52)	19.99	(1.76)
Marche	1532.5	15 321	(229)	28.71	(0.79)	29.85	(0.59)	15.24	(1.22)
Lazio	5399.0	14 973	(299)	32.92	(0.98)	26.77	(0.59)	17.90	(1.16)
Abruzzo	1307.6	13 264	(397)	30.29	(1.10)	31.33	(0.90)	22.30	(1.92)
Molise	320.5	11 844	(516)	31.19	(1.30)	28.15	(0.70)	27.88	(2.62)
South									
Campania	5790.6	10 912	(251)	34.16	(0.91)	26.98	(0.63)	34.12	(1.88)
Puglia	4070.7	10 318	(196)	31.95	(1.02)	29.60	(0.68)	35.74	(2.22)
Basilicata	592.7	11 367	(281)	29.39	(1.07)	31.70	(0.72)	30.81	(3.05)
Calabria	2001.2	11 090	(141)	31.98	(1.19)	27.77	(0.87)	32.97	(2.41)
Islands									
Sicilia	5017.0	10 147	(287)	34.27	(0.80)	26.59	(0.57)	39.79	(1.57)
Sardegna	1657.6	13 617	(404)	30.45	(1.50)	29.38	(0.99)	19.04	(2.32)

Table 5: Estimates of the Spanish regional indicators of income level and inequality.

Region	*Population size* (in '000)	*Median income*		*Gini coefficient*		$L(0.5)$		*Poverty rate*	
		Estimate	St.error	Estimate	St.error	Estimate	St.error	Estimate	St.error
North									
Galicia	2721.2	10 097	(285)	30.44	(1.03)	29.04	(0.71)	23.64	(2.09)
Asturias	1058.2	12 080	(352)	31.87	(1.82)	28.85	(0.99)	14.48	(1.83)
Cantabria	560.4	12 174	(373)	32.11	(1.74)	28.31	(1.18)	11.75	(1.97)
Pais Vasco	2118.6	13 914	(277)	27.50	(1.04)	31.02	(0.70)	11.81	(1.69)
Navarra	592.3	15 923	(530)	29.09	(1.14)	29.42	(0.92)	10.77	(2.05)
La Rioja	303.5	11 051	(432)	27.73	(1.12)	30.37	(0.72)	22.40	(2.46)
Aragon	1267.4	11 305	(411)	30.32	(1.08)	28.83	(0.71)	16.16	(1.61)
Centre									
Madrid	5995.5	12 863	(421)	30.80	(0.90)	28.41	(0.61)	12.00	(1.15)
Castilla y Leon	2481.6	9120	(220)	32.42	(0.90)	27.61	(0.55)	30.74	(1.87)
C'lla-La Mancha	1911.3	9161	(322)	31.46	(1.07)	28.09	(0.72)	32.73	(1.86)
Extremadura	1072.9	8531	(289)	31.21	(1.31)	28.50	(0.79)	29.00	(2.84)
East									
Catalunya	7010.7	13 577	(231)	27.96	(0.59)	30.22	(0.47)	12.29	(1.30)
Valencia	4700.3	11 406	(171)	32.95	(1.67)	27.98	(0.95)	17.12	(1.83)
Balears	1000.0	13 286	(382)	29.38	(1.16)	29.87	(0.80)	12.59	(1.77)
South									
Andalucia	7855.8	9607	(188)	32.59	(0.75)	27.58	(0.49)	28.34	(1.65)
Murcia	1353.1	10 007	(284)	29.91	(1.18)	28.87	(0.72)	26.38	(2.52)
Ceuta	71.5	10 401	(1089)	35.99	(1.86)	24.02	(1.40)	30.08	(3.44)
Melilla	67.0	10 274	(1127)	35.96	(2.08)	24.69	(1.67)	26.64	(4.71)
Canarias	1975.2	9766	(318)	33.69	(1.18)	26.02	(0.89)	30.76	(2.60)

Table 6: Estimates of the regional indicators of income level and inequality in Czech Republic.

Region	*Population size* (in '000)	*Median income*		*Gini coefficient*		$L(0.5)$		*Poverty rate*	
		Estimate	St.err.	Estimate	St.err.	Estimate	St.err.	Estimate	St.err.
Praha	1184.9	5378	(103)	31.63	(1.79)	28.18	(1.00)	3.77	(1.83)
Strední Čechy	1166.7	4811	(85)	29.10	(1.32)	29.95	(0.74)	7.11	(1.57)
Jihozápad	1181.9	5030	(71)	20.76	(0.55)	35.36	(0.43)	3.74	(1.20)
Severozápad	1127.7	4360	(46)	27.47	(1.78)	31.46	(0.94)	12.13	(1.59)
Severovýchod	1485.8	4485	(57)	22.65	(0.97)	34.17	(0.62)	6.18	(1.50)
Jihovýchod	1642.7	4418	(53)	22.73	(0.71)	34.33	(0.42)	7.04	(1.58)
Střední Morava	1229.5	4347	(68)	25.42	(0.97)	31.91	(0.57)	10.18	(1.90)
Moravskoslezsko	1250.0	4144	(59)	25.17	(0.71)	31.38	(0.53)	15.18	(1.52)

References

Anand, S., and S. M. Kanbur (1993). 'The Kuznets Process and the Inequality-Development Relationship', *Journal of Development Economics*, 40, 25–72.

Barro, R. and Sala-i-Martin , (1991), 'Convergence across States and Regions', *Brooking Papers on Economic Activity*, 22, 107–182.

Beckfield, J. (2009). 'Stratifying Europe: Regional Integration Raises Income Inequality but Brings Convergence in the EU', *Harvard University Multidisciplinary Program on Inequality and Social Policy*.

Biewen, M. (2002). 'Bootstrap inference for inequality, mobility and poverty measurement', *Journal of Econometrics*, 108, 317–42.

De Rynck, S., McAleavey P. (2001). 'The Cohesion Deficit in Structural Fund Policy', *Journal of European Public Policy*, Vol. 8, 541–557.

Eurostat (2007), *Eurostat regional yearbook 2007*, Eurostat Statistical books, European Commission.

Förster, M.F., Jesuit D., and Smeeding T. (2005). 'Regional Poverty and Income Inequality in Central and Eastern Europe: Evidence from the Luxembourg Income Study', in: R. Kanbur and A.J. Venables (eds.), *Spatial Inequality and Development*, pp. 311-347. Oxford University Press.

Hoffmeister, O. (2009). 'The spatial structure of income inequality in the enlarged EU'. *Review of Income and Wealth*, 55, 101–127.

Jesuit, D. K., Rainwater L., Smeeding T. (2003). 'Regional poverty within rich countries'. In J.A. Bishop, Y. Amiel (eds) *Inequality, Welfare and Poverty: Theory and Measurement*, Elsevier Science, New York, 345–377.

Jesuit D. K. (2008). 'Subnational analyses using the Luxembourg Income Study (LIS) Data Archive'. In G. Grozinger, K.S. Spies, W. Matiaske (eds), *Europe and Its Regions: The Usage of European Regionalized Social Science Data*, Cambridge Scholars Press, Cambridge, UK, 57–65.

Kuznets, S. (1955). 'Economic growth and income inequality', *American Economic Review*, 45, 1–28.

Le Gallo, J. (2004). 'Space-time analysis of GDP disparities among European regions', *International Regional Science Review*, 27, 138–163.

Li, H., L. Squire, and H. Zou (1998). 'Explaining International and Intertemporal Variations in Income Inequality', *Economic Journal*, 108, pp. 26–43.

Longford, N. T. (2005). *Missing Data and Small-area Estimation. Analytical Equipment for the Survey Statistician.* Springer-Verlag, New York.

Mahler, V. A. (2002). 'Exploring the subnational dimension of income inequality: an analysis of the relationship between inequality and electoral turnout in the developed countries', *International Studies Quarterly*, 46, 117–142.

Mills, J. A. and Zandvakili, S. (1997). 'Statistical Inference via Bootstrapping for Measures of Inequality', *Journal of Applied Econometrics*, 12, 133–150.

Perugini, C., Martino G. (2008). 'Income inequality within European regions: determinants and effects on growth', *Review of Income and Wealth*, 54, 373–406.

Pittau, M. G., Zelli R. (2006). 'Empirical Evidence of Income Dynamic Across EU Regions', *Journal of Applied Econometrics*, 21, 605–628.

Potthoff, R.F., K.G. Manton and M.A. Woodbury (1993). 'Correcting for nonavailability bias in surveys by weighting based on number of callbacks', *Journal of the American Statistical Association*, 88.

Quah, D. (1996). 'Regional convergence clusters in Europe', *European Economic Review*, 35, 951–958.

Stewart, K. (2002). 'Measuring Well-Being and Exclusion in Europe's Regions'. *Luxembourg Income Study Working Paper No. 303.*

Weil, D. N. (2005), *Economic Growth.* Pearson Education.