Departament d'Economia Aplicada

A sensitivity analysis of poverty definitions

Nicholas T. Longford
Catia Nicodemo

DOCUMENT DE TREBALL

09.08

Universitat Autònoma de Barcelona

Facultat de Ciències Econòmiques i Empresarials

A sensitivity analysis of poverty definitions

Nicholas T. Longford

SNTL and Universitat Pompeu Fabra, Barcelona, Spain,[*]

Catia Nicodemo

Universitat Autònoma de Barcelona, Campus Bellaterra, Cerdanyola, Spain, and IZA Bonn, Germany

Abstract

We conduct a sensitivity analysis of several estimators related to household income, to explore how some details of the definitions of the variables concerned influence the values of the common estimates, such as the mean, median and (poverty) rates. The purpose of this study is to highlight that some of the operational definitions entail an element of arbitrariness which leaves an undesirable stamp on the inferences made. The analyses use both a cross-sectional and a longitudinal (panel) component of the EU-SILC database.

Keywords: *EU-SILC database, poverty rate, sampling weights.*

---

[*]Correspondence: N. T. Longford, SNTL and Departament d'Economia i Empresa, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain; email: NTL@sntl.co.uk.

1 Introduction

Poverty or deprivation, of an individual or household, is a general and multifaceted concept that defies a straightforward definition or measurement and elicitation. The design of every large-scale survey of poverty faces the conflicting demands of collecting more detail from the subject and keeping the response burden to minimum, so as to encourage full cooperation. Definitions of poverty range from conceptual, such as having insufficient resources for the acquisition and maintenance of items and services regarded as indispensable for everyday existence, to practical, such as comparing the household income against a specified standard (the poverty threshold). For a conceptual definition of poverty, see Sen (1992). The values of variables related to the practical definitions can be elicited by questionnaire instruments in surveys, or, at least in principle, extracted from population registers. The variables related to conceptual definitions are invariably latent, and have to be connected to their manifest versions by models that are often difficult or impossible to verify. Further difficulties arise when some subjects' responses involve errors due to imperfect recall or intentional misreporting.

The study of poverty is frequently reduced to the study of income (of an adult individual or a household), and poverty status is defined coarsely as a dichotomous variable (being poor or not); Atkinson *et al.* (2002). Resources acquired, such as income, are not spent immediately, and demand for them may be uneven. Borrowing funds can help tie a household over a period of high demand (purchase of expensive long-lasting goods or a chance accumulation of several expenses), and so the level of income, whichever way it may be adjusted for household composition, is at best a crude measure of poverty.

Several papers study the poverty in a single EU country or its part; Delhausse, Lüttgens and Perelman (1993) in Wallonia, Belgium, Lovillier and Verger (1997) in France, and Pérez-Mayo (2004) in Spain. The European Survey of Income and Living Conditions (EU-SILC), Central Statistical Office (2005), is an important source of data about the income in the countries of the European Union. It is funded by the European Commission and regulated by the European Parliament and the Council. These institutions have a vested interest in social protection and they promote policies conducive to redistribution of wealth. The Survey is conducted since 2003 annually.

In 2003 it had 12 and in 2005 and 2006 26 participating countries. The designs and protocols and the questionnaire instruments in the national versions of EU-SILC are harmonized (made as similar as possible), enabling credible international comparisons. Lelkes and Zólyomi (2008) explore the poverty rates in the countries that participate in EU-SILC.

The survey has a cross-sectional and a longitudinal (panel) element. At the time of writing, for most countries, panel data were available only for 2005 and 2006, and for a few countries also for 2004. Although EU-SILC collects information about housing condition, health, education and employment, our study is concerned only with income, and we analyse data only from the cross-sectional survey in 2007 and data from the panel $2005-2006$ for all the participating countries.

We are concerned with aspects of definitions of poverty that have an element of arbitrariness. For example, according to a commonly used definition, a household is classified as poor if its equivalised income in the studied year is lower than 60% of the national median equivalised household income. We refer to this cut-off value as the *poverty threshold*. In this example, the poverty threshold is country-specific but, in principle, a single (universal) threshold could also be defined. With such a threshold, a vast majority of the population in East European countries would be classified as poor, for example, between 80 and 92 per cent in the Baltic states and Slovakia (Pichaud, 2000). Country-specific thresholds are more realistic, even if they distort the comparisons intended for poverty that refers to an absolute level.

The cut-off level of 60% is chosen by convention, as is the way how the total income is equivalised — adjusted for household composition: one adult member of the household is counted as 1.0, every other member over 14 years of age as 0.5, and every child (member aged 14 or younger) as 0.3. For example, a household comprising three members older and two members younger than 14 years of age has equivalised household size $1 + 2 \times 0.5 + 2 \times 0.3 = 2.6$. If the total income of the household (in a year) in $52\,000$ Euro, then its equivalised household income is $52\,000/2.6 = 20\,000$ Euro. Although the factors 0.5 and 0.3 are reasonable, it might be difficult to argue against some similar factors, such as 0.6 and 0.4. Even the composition of income entails some ambiguities, foremost among them whether the gross or net income should be recorded, and whether some payments (goods or services received) in kind should be taken into

account.

We conduct a number of sensitivity analyses which explore the impact of altering some aspects of the definitions of (national and regional) average income and poverty rate on the comparisons of the countries or regions. We summarise the sensitivity by diagrams in which the estimates of the quantities of interest are plotted as functions of the adjustment.

In the next section, we use the following variables recorded in the EU-SILC database:

1. *Household identification* (`HB030`)

2. *Personal identification* (`PB030`)

3. *Household type* (`HX060`)

4. *Total household disposable income* (`HY020`)

5. *Total household gross income* (`HY010`), when `HY020` is not recorded

6. *Household cross-sectional weight* (`DB090`)

7. *Longitudinal weight* (for two years of duration) (`RB062`)

8. *Region* (`DB040`)

9. *Year of birth* (`RB080`)

See Eurostat (2009) for details. Year of birth and Personal identification are recorded for individuals; the other variables are recorded for households. The database contains equivalised household income and equivalised household size, but in some analyses we calculate these by rules that differ from the operational.

The next section deals with estimation of the median income with a range of definitions of the equalised household size (equivalisations). The median income is estimated for countries and regions of four countries for which at least six regions are recorded in the database.

Section 3 studies the sensitivity of the poverty rate (percentage of individuals in poverty) with respect to the poverty threshold and Section 4 applies this approach to subpopulations of specific interest: single-parent households and children. Section

5 studies the sensitivity of the mean poverty gap, defined as the mean shortfall of equivalised income with respect to the threshold. Section 6 explores a continuum of estimators that use trimming which have the (weighted) sample mean and the median as their extremes. Section 7 is concerned with estimation of the rates of transition to and from poverty and the rates of remaining poor.

2   Median income and equivalised household size

In this section, we study how the comparisons of the median equivalised income of countries are affected by the definition of the equivalised household size. Suppose a household has H_1 adult and H_2 under-age members. Then according to the definition of the equivalised household size implemented in EU-SILC, the household's equivalised size is $1 + 0.5(H_1 - 1) + 0.3H_2$. As general alternatives, we consider the equivalisations

$$1 + (0.4 + 0.005K)(H_1 - 1) + (0.25 + 0.0025K)H_2, \qquad K = 0, 1, \ldots, 80, \quad (1)$$

$$1 + (0.45 + 0.004K)(H_1 - 1) + 0.3H_2, \qquad K = 0, 1, \ldots, 75. \quad (2)$$

so that the factors are in the range (0.4, 0.8) for $H_1 - 1$ and (0.25, 0.45) for H_2 in the first set equivalisations. In the second set we alter only the factor of $H_1 - 1$; it ranges from 0.45 to 0.75. In the first set, the operational equivalisation corresponds to $K = 20$, and in the second to $K = 12.5$. We refer to K as *offset*. No generality is lost by associating the first adult member (head) of a household with unity ($H_0 = 1$), because $H_0 \neq 1$ can be compensated by an adjustment of H_1 and H_2.

We apply these equivalisations, evaluate the corresponding equivalised household income for each household in the 2007 national surveys and estimate the national median equivalised household income. With increasing offset each equivalised income is reduced or unaltered, but we are concerned only with comparisons of the national and regional means. Sensitivity is broadly defined as a substantial change in the (estimated or exact) summary as a result of a small change in the factor; sensitivity is an undesirable property. It cannot be clinically evaluated unless we specify what we mean by 'small' and 'substantial'. In our context, this may be an alteration of the definition of the equivalisation that could equally well have been adopted, except for the preference for rounded factors of $H_1 - 1$ and H_2.
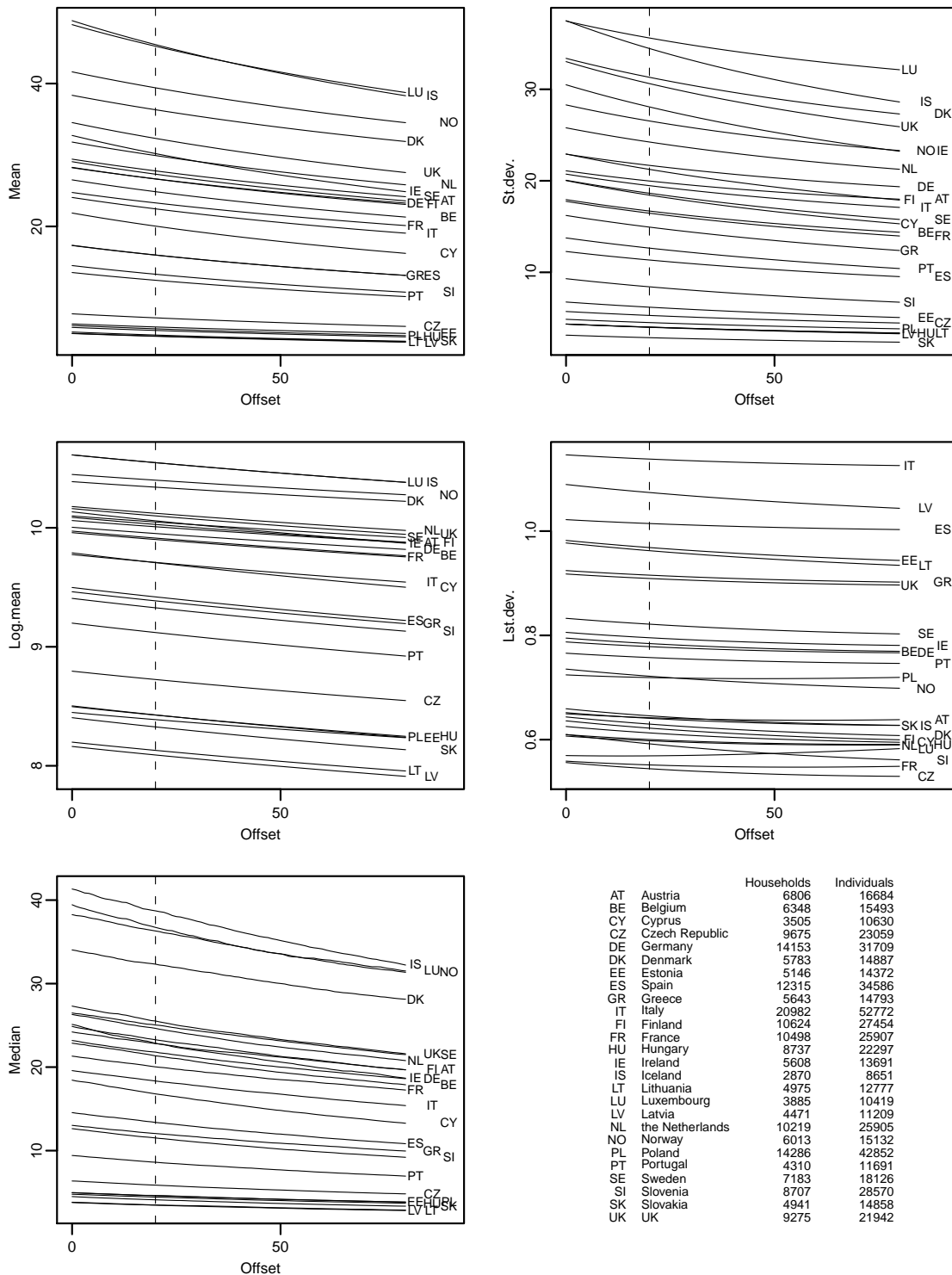
Figure 1: Sensitivity of the estimates of the national mean, log-mean and median income with respect to the equivalisation in (1). The means (and the corresponding standard deviations) and medians are in thousands of Euros, the log-means are in log-Euros.

6

The sensitivity of the national comparisons of mean, log-mean and median income with respect to altering the factors of $H_1 - 1$ and H_2 is summarised in the left-hand panels of Figure 1. The top two right-hand panels summarise the estimates of the standard deviations on the linear and log scales and the bottom panel lists the countries and the numbers of households and individuals in their samples in EU-SILC 2007. The acronyms (AT, BE, ..., UK) are used in the margins of the panels. The vertical dashes indicate the operational definition, $K = 20$, with factors $(1, 0.5, 0.3)$.

All the population summaries take into account the sampling weights. The mean of a variable Y is estimated from the sample values $(y_1, \ldots, y_n)$ with sampling weights $(w_1, \ldots, w_n)$ by the ratio estimator

$$\hat{\mu} = \frac{1}{w_+} \sum_{i=1}^{n} w_i \, y_i \,,$$

and the population variance by

$$\hat{\sigma}^2 = \frac{1}{w_+} \sum_{i=1}^{n} w_i \, (y_i - \hat{\mu})^2 \,.$$

The standard deviation is estimated by the square root $\sqrt{\hat{\sigma}^2}$. Quantile $q \in (0, 1)$ of Y is estimated by the sample quantile, evaluated by the following algorithm. Let $y^{(1)} \leq \ldots \leq y^{(n)}$ be the values of the equivalised household income sorted into ascending order. Let $w^{(1)}, \ldots, w^{(n)}$ be the corresponding (permuted) sampling weights and $u^{(1)} = w^{(1)}$, $u^{(2)} = w^{(1)} + w^{(2)}$, ..., $u^{(n)} = w^{(1)} + \cdots + w^{(n)} = w_+$ their cumulative totals. The sample q-quantile of Y is defined as the value $y^{(k)}$ such that $u^{(k-1)} < qw_+ < u^{(k)}$. If $qu^{(n)} = q^{(k)}$ for some k, then sample quantile is set to $(y^{(k)} + y^{(k+1)})/2$.

In Section 6, we evaluate the sample (weighted) trimmed mean. For $q \in (0, 1)$, the sample q-trimmed mean can be defined as a generalisation of the sample mean and sample median. Suppose k and h are such that $u^{(k-1)} < qw_+ < u^{(k)}$ and $u^{(h)} < (1 - q)w_+ < u^{(h+1)}$. Then the sample q-trimmed mean is defined as

$$\frac{1}{1 - 2qw_+} \left[\left(u^{(k)} - qw_+ \right) y^{(k)} + \sum_{i=k+1}^{h} w^{(i)} y^{(i)} + \left\{ u^{(h+1)} - (1 - q)w_+ \right\} y^{(h+1)} \right] \,.$$

That is, observations are discarded at either extreme up to the fraction q of the total weight (qw_+), and the weighted sample mean is calculated for the remainder. The observations at the q and $1 - q$ quantile contribute to the weighted sample mean only
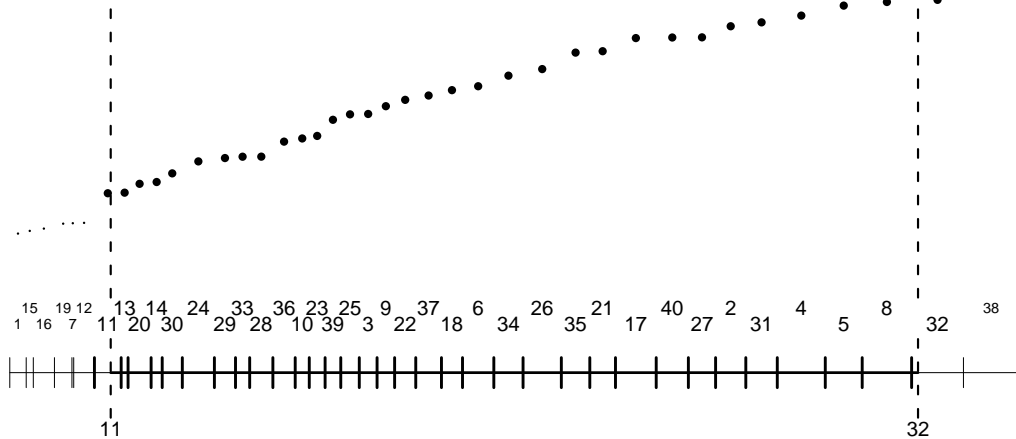
Figure 2: The q-trimmed mean. An illustration with $n = 40$ observations and $q = 0.1$.

partially. An illustration is given in Figure 2. The observations, marked by dots, are presented in the ascending order. The height at which the dot is printed represents the value of y. The order numbers are printed underneath in two rows, to avoid overprinting. For the q fraction of the observations at either extreme, the dots are small and the order numbers are printed by a smaller font. The weights are presented by the horizontal segment. The weights of the extreme observations correspond to the thin part of the segment. Observations 11 and 32 contribute to the 0.1-trimmed mean only by fractions of their respective weights. The weighted sample mean coincides with $q = 0$ trimmed mean (no trimming) and the sample median with $q = 0.5$ trimmed mean (full trimming).

The diagrams of the means, log-means and median confirm that income tends to be highest in the northwestern and lowest in the eastern European countries. The choice of the scale (linear or log) and using the mean or median, can itself be regarded as an issue of sensitivity. In all three left-hand panels, Luxembourg, Iceland, Norway and Denmark have the highest values of the summary, but their relative distances differ. For example, Luxembourg and Iceland have nearly identical values for the mean and log-mean for all offsets, but for the median they differ appreciably throughout the range $K \in (0, 80)$. The distance of these four countries from the countries next in the order is relatively greater for the median.

8

The eastern European countries have the lowest summaries, and they differ very little for the mean and median. However, the estimate of the log-mean for Czech Republic is greater and for Latvia and Lithuania smaller than for the other four countries.

Total lack sensitivity corresponds to curves that are parallel or concentric, or would become parallel by a monotone transformation. In particular, such curves do not intersect (or intersect at a single point), and maintain their relative distances. The summaries for mean, log-mean and median are quite insensitive. The curves cross in a few instances (e.g., Luxembourg and Iceland for the mean; Italy and Cyprus for the log-mean), or they converge, so that they are distant for one offset and close for another (e.g., Estonia, Poland and Hungary for the log-mean; Ireland and several countries, including Germany, for the median). However, for a small change in the factors we have a small changes in the relative magnitudes of the summaries.

The estimates of the standard deviations (right-hand panels of Figure 1) are somewhat more sensitive to the setting of the factors, more so on the linear scale. On the log scale, Luxembourg stands out; its curve increases throughout the range of offset quite steeply, whereas for most other countries it is flat or decreasing at a slow rate.

The standard deviation curves for the means closely resemble the curves for the mean. To a large extent this is due to the log-normal nature of income. The log-normal distribution, derived by exponentiating the normal distribution $\mathcal{N}(\mu, \sigma^2)$, has expectation $\exp(\mu + \frac{1}{2}\sigma^2)$ and variance $\exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}$, so the mean and variance of log-normally distributed variables are strongly positively associated even when the underlying normals have equal variances. This is borne out by the panel for the standard deviations on the log scale, which has no apparent association with the panel for the log-means.

The sensitivity of the five summaries of income for the equivalisation given by (2) is summarised in Figure 3. The conclusions are very similar to those based on Figure 1. The two sets of equivalisations differ only slightly because in the samples for most countries there are far fewer children than adults.

2.1   Income in the regions

We study the sensitivity of the summaries of income for regions in four countries, Czech Republic, Poland, Spain and France, in each of which a division into a relatively large
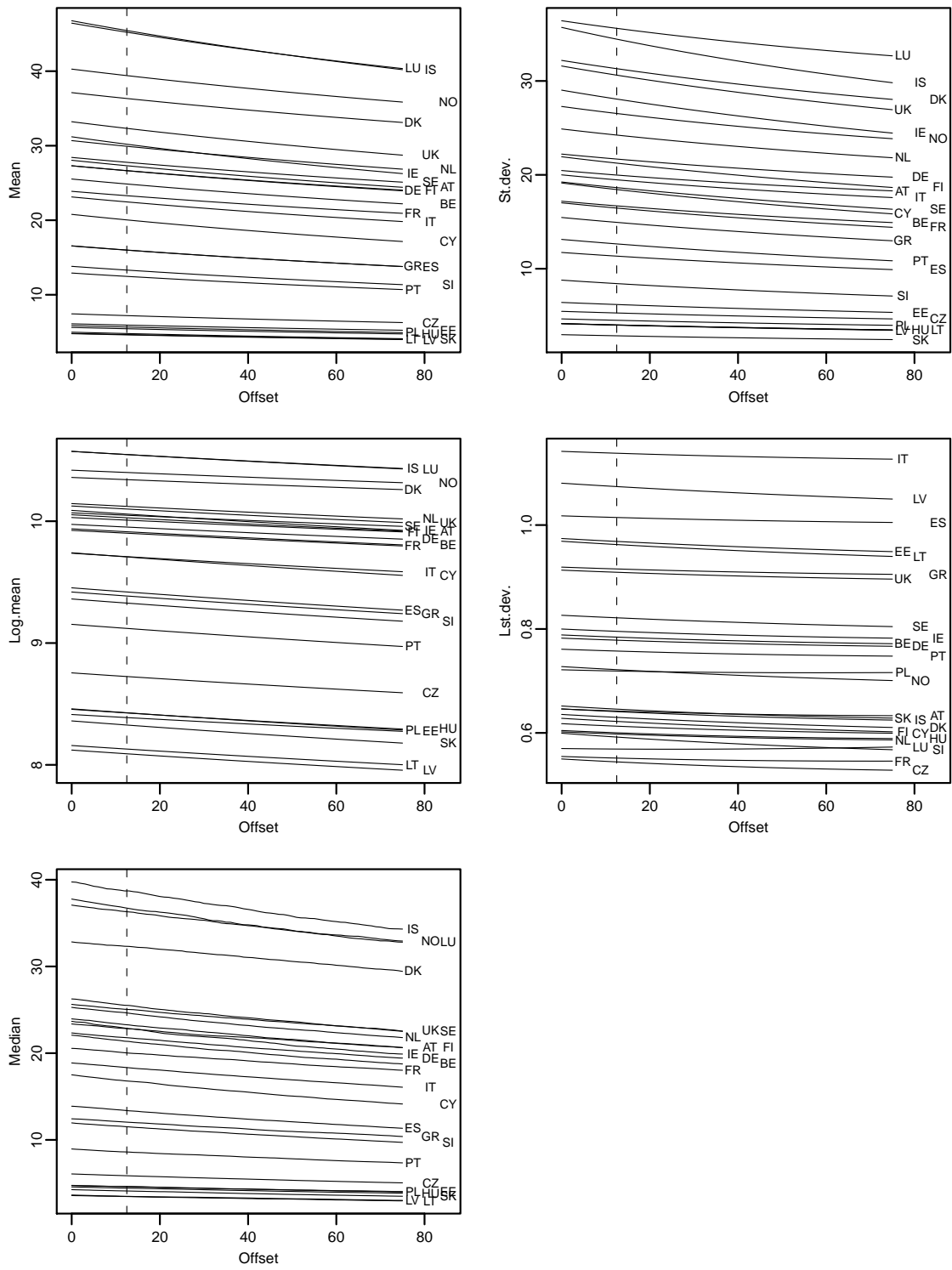
Figure 3: Sensitivity of the estimates of the national mean, log-mean and median income with respect to the equivalisation in (2).

number of regions is defined. In the other countries, no regions are identified, or they are aggregated to define fewer than six regions.

Figures 4 – 7 display the sensitivity curves for the regions of the four countries with respect to the equivalisation given by (1). The sensitivity curves for the country are drawn by thick dashes, and the regions are listed in the bottom right-hand panel, together with their sample sizes.

The diagrams for Czech Republic and Poland are difficult to assess because these countries have relatively few regions in the data, eight and six, respectively. The estimates of the mean, log-mean and median are close to being parallel, although they intersect for regions 5 and 6 (North-East and South-East) and 7 and 8 (Central Moravia and Moravian Silesia) of the Czech Republic. The summaries for the eight regions are in the same order in all three left-hand panels, with the capital Prague (region 1) standing out as having the highest mean household income. The curves for the median are not smooth because the median is a discontinuous function of the data. The estimates of the standard deviations are much more sensitive to equivalisation. The shape of the curve for region 8 differs substantially from the curves for the other regions, especially for log-mean. The estimated curves for the mean and log-mean are in some discord — the two regions with highest estimates swap their order for all offsets, and regions 3, 5 – 8 have similar standard deviations on the linear scale (mean), but differ substantially on the log scale (log-mean).

The sensitivity curves for the regions of Poland are quite insensitive to the details of equivalisation, but they are highly sensitive to the choice of the estimator (scale). Region 1 has by far the highest estimated mean, but its estimated curves for the log-mean and median do not stand out as much. The ranks of the estimated standard deviations on the linear and log scales differ substantially.

The sensitivity curves for Spain and France are very close to being parallel for the estimates of the mean, log-mean and median. The curves for the standard deviations on the linear scale intersect in many instances, but are close to being parallel. The curves on the log scale show considerable sensitivity for the French regions and hardly any for the Spanish regions. The choice of the scale is important for the standard deviation, but much less so for location (mean, log-mean or median).
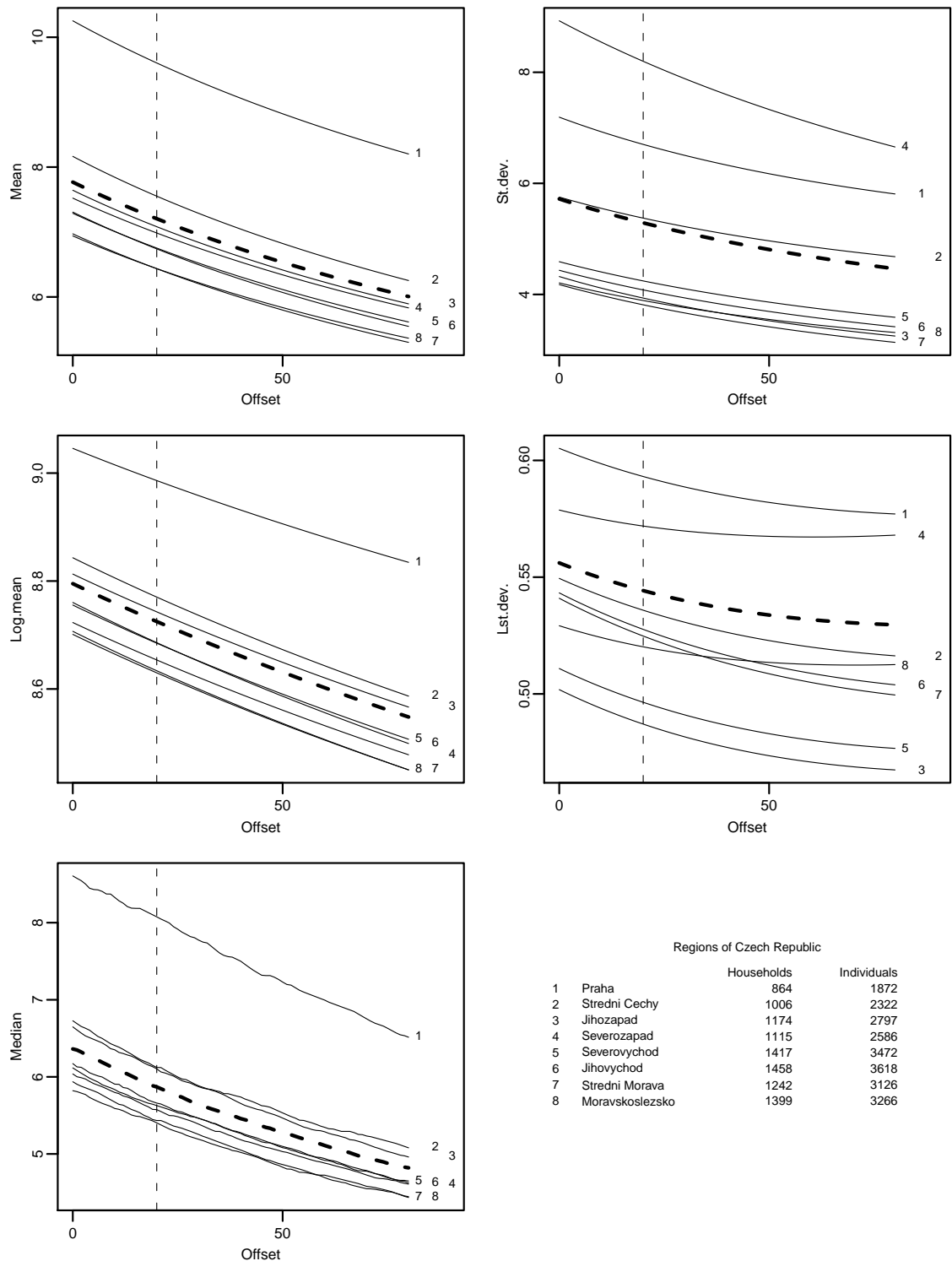
11

Figure 4: Sensitivity of the estimates of the national mean, log-mean and median income with respect to the equivalisations in (1) for the regions of Czech Republic.
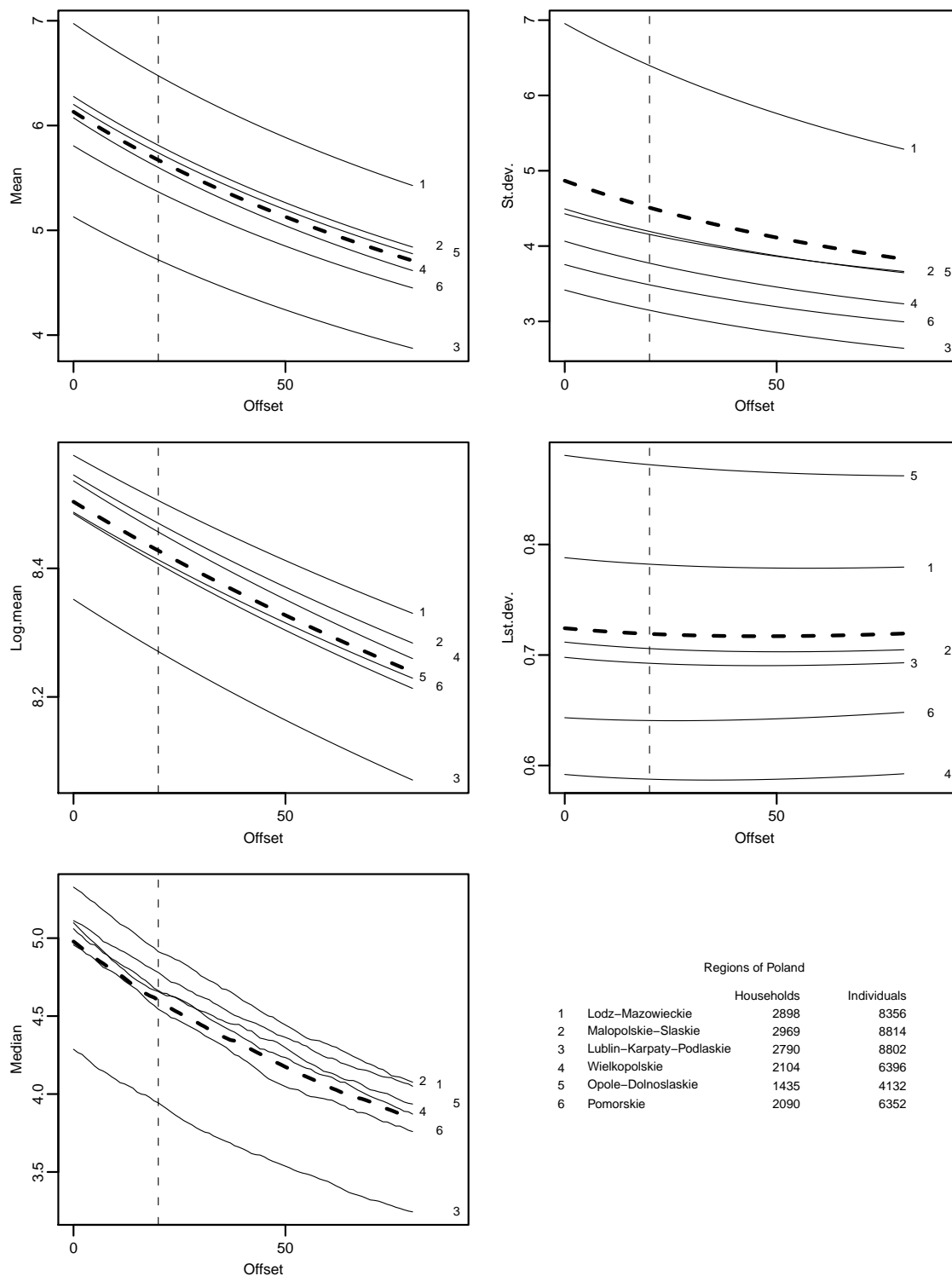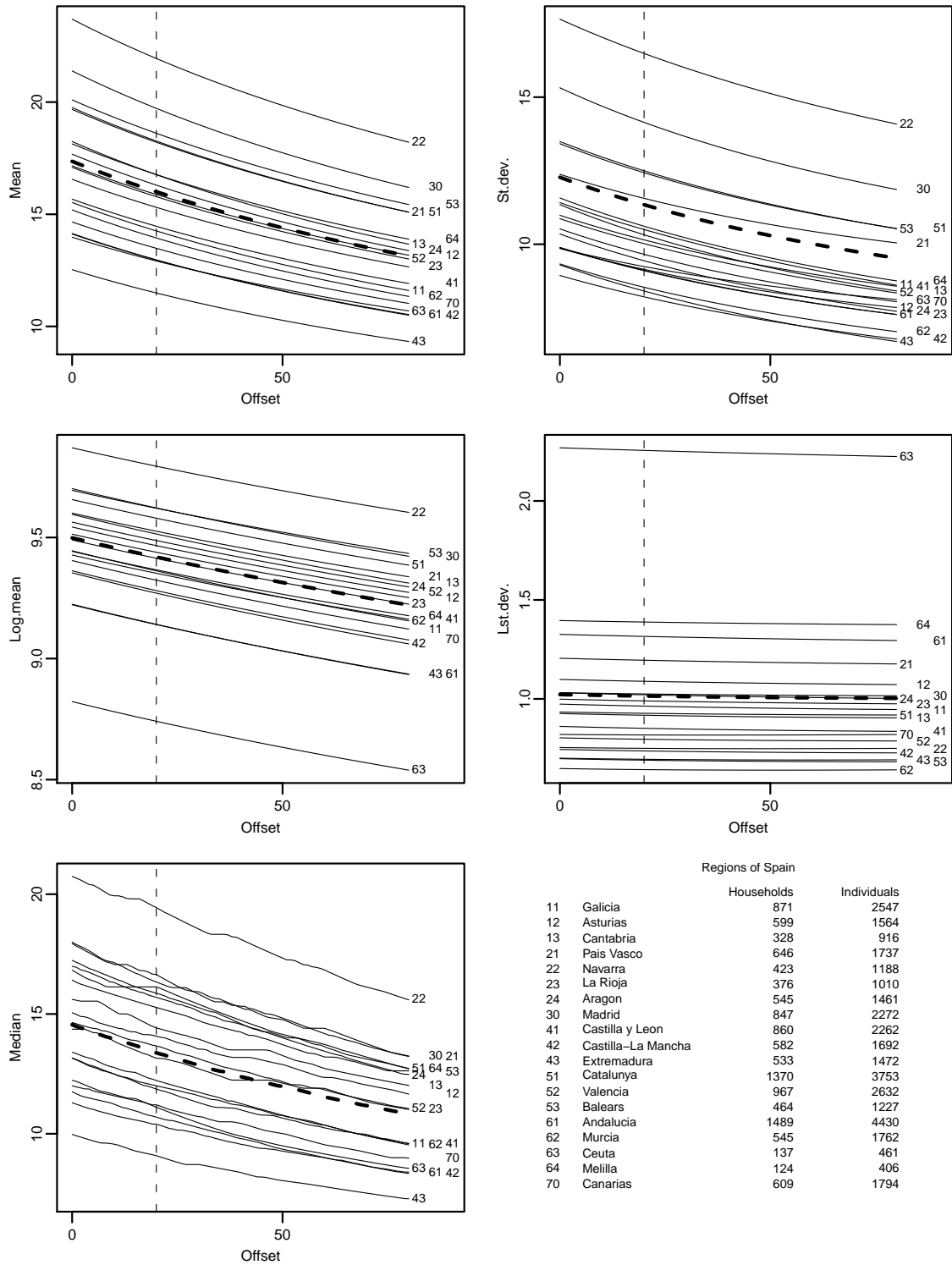
Mean

St.dev.

Offset

Offset

Log.mean

Lst.dev.

Offset

Offset

Median

Offset

Regions of Poland

		Households	Individuals
1	Lodz–Mazowieckie	2898	8356
2	Malopolskie–Slaskie	2969	8814
3	Lublin–Karpaty–Podlaskie	2790	8802
4	Wielkopolskie	2104	6396
5	Opole–Dolnoslaskie	1435	4132
6	Pomorskie	2090	6352

Figure 5: Sensitivity of the estimates of the national mean, log-mean and median income with respect to the equivalisations in (1) for the regions of Poland.

13

Figure 6: Sensitivity of the estimates of the national mean, log-mean and median income with respect to the equivalisations in (1) for the regions of Spain.
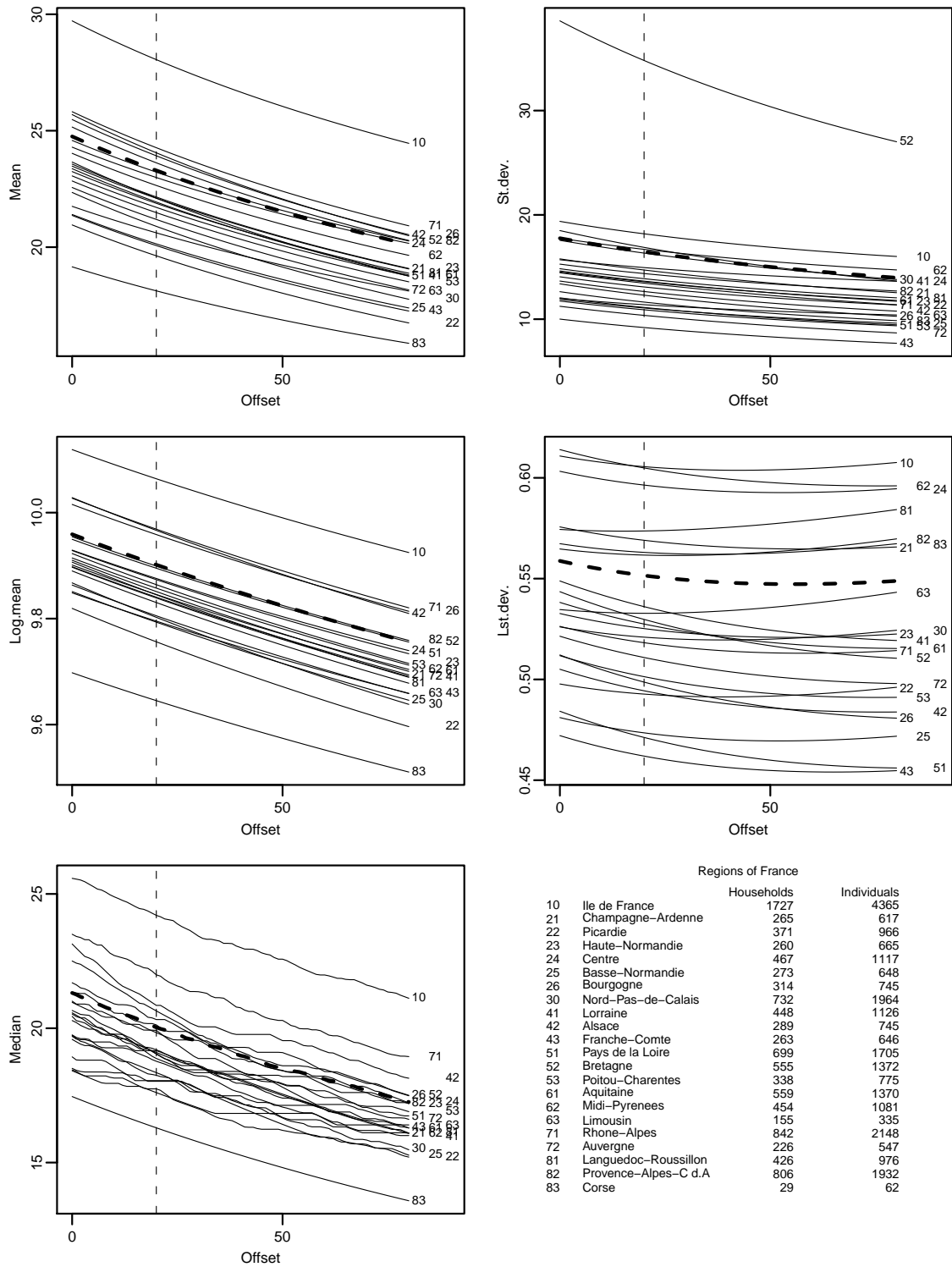
Figure 7: Sensitivity of the estimates of the national mean, log-mean and median income with respect to the equivalisations in (1) for the regions of France.

15

3 Poverty rate and the threshold

The poverty status, that is, being poor or not, is commonly defined by relating the household equivalised income to a standard, or threshold, set to a given percentage of the national median equivalised household income. This percentage is set by convention to 60%, but there is no profound rationale for this choice. The poverty rate is defined as the percentage of individuals who are classified as poor. As an alternative, the poverty rate could be defined as the percentage of households that have equivalised income lower than the set threshold. We explore how sensitive are the comparisons of the estimated poverty rates with respect to the percentage that defines the threshold and whether individuals or households are counted. We refer to these two kinds of poverty rates as individual- and household-level rates.

The country's estimated individual-level poverty rates as functions of the threshold (standard) are plotted in Figure 8. The diagram shows that the comparisons of the countries are sensitive to the setting of the threshold. For example, Norway has a relatively low poverty rate with respect to the threshold of 80% (24.9%, the fifth lowest), and relatively much higher rate with respect to the threshold of 40% (5.1%, the ninth highest). Ireland has a relatively low poverty rate for 40% threshold, 3.0%, but the seventh highest poverty rate for 80%, equal to 32.1%. The ranks of the poverty rates for several other countries differ substantially between the two extreme settings of the threshold, although some countries have the highest poverty rates for all thresholds (Greece, Italy, Poland and Spain) and some the lowest (Czech Republic).

Figure 9 displays the sensitivity curves for the estimates of the household-level poverty rates. The diagram closely resembles its counterpart for individual-level in Figure 8, although there are some notable differences. (The two diagrams have the same vertical scale.) For example, the estimated household-level poverty rates in Slovakia are the second lowest (following Czech Republic) throughout the range of the threshold percentages $40-80\%$, and the rates for Latvia are uniformly the highest. The estimated individual-level rates for Poland are among the highest throughout the range, but their household-level counterparts are near the average throughout. There are substantial differences between the relative positions of the household- and individual-level poverty rates for several other countries.
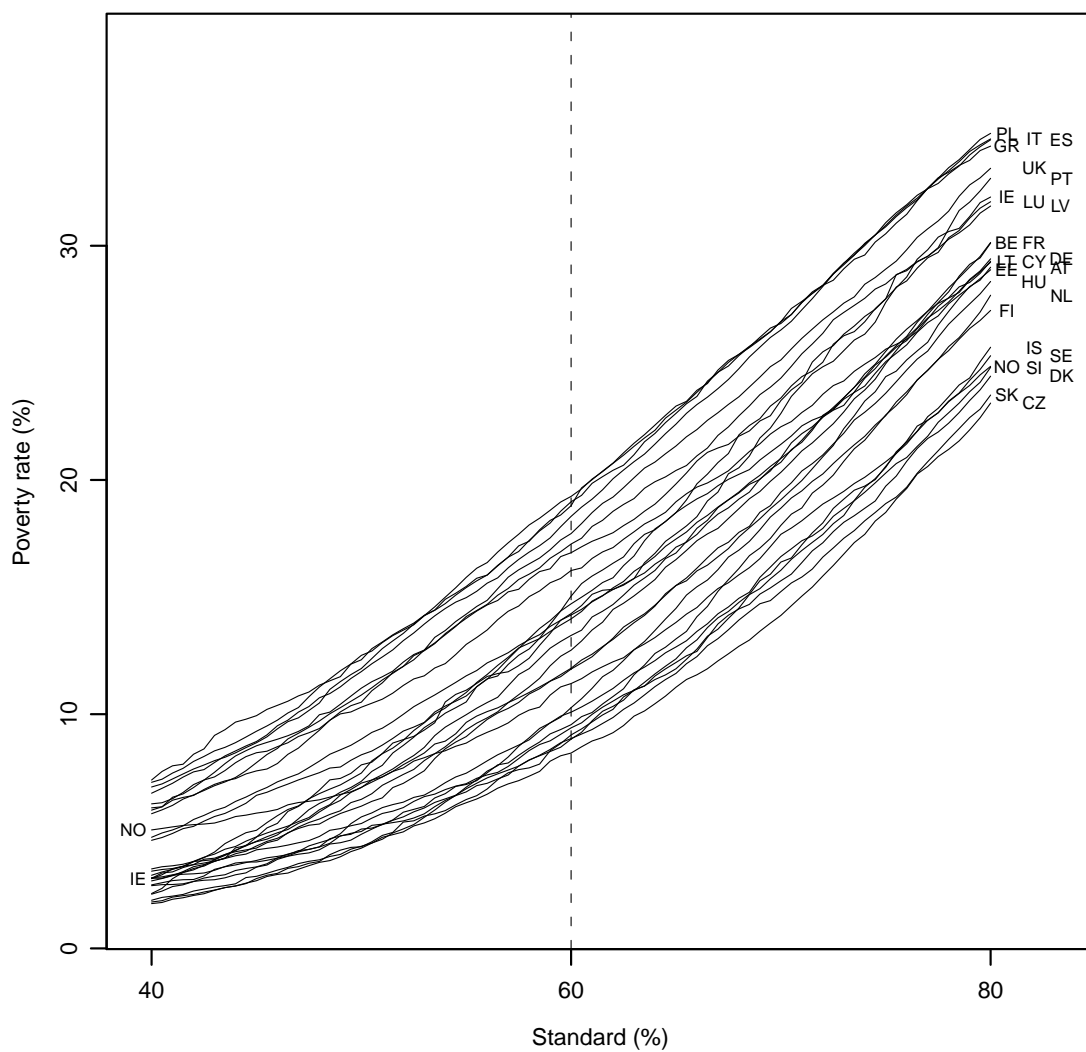
Figure 8: Sensitivity of the estimates of the individual-level poverty rate with respect to the threshold of $40-80\%$ of the median equivalised household income.
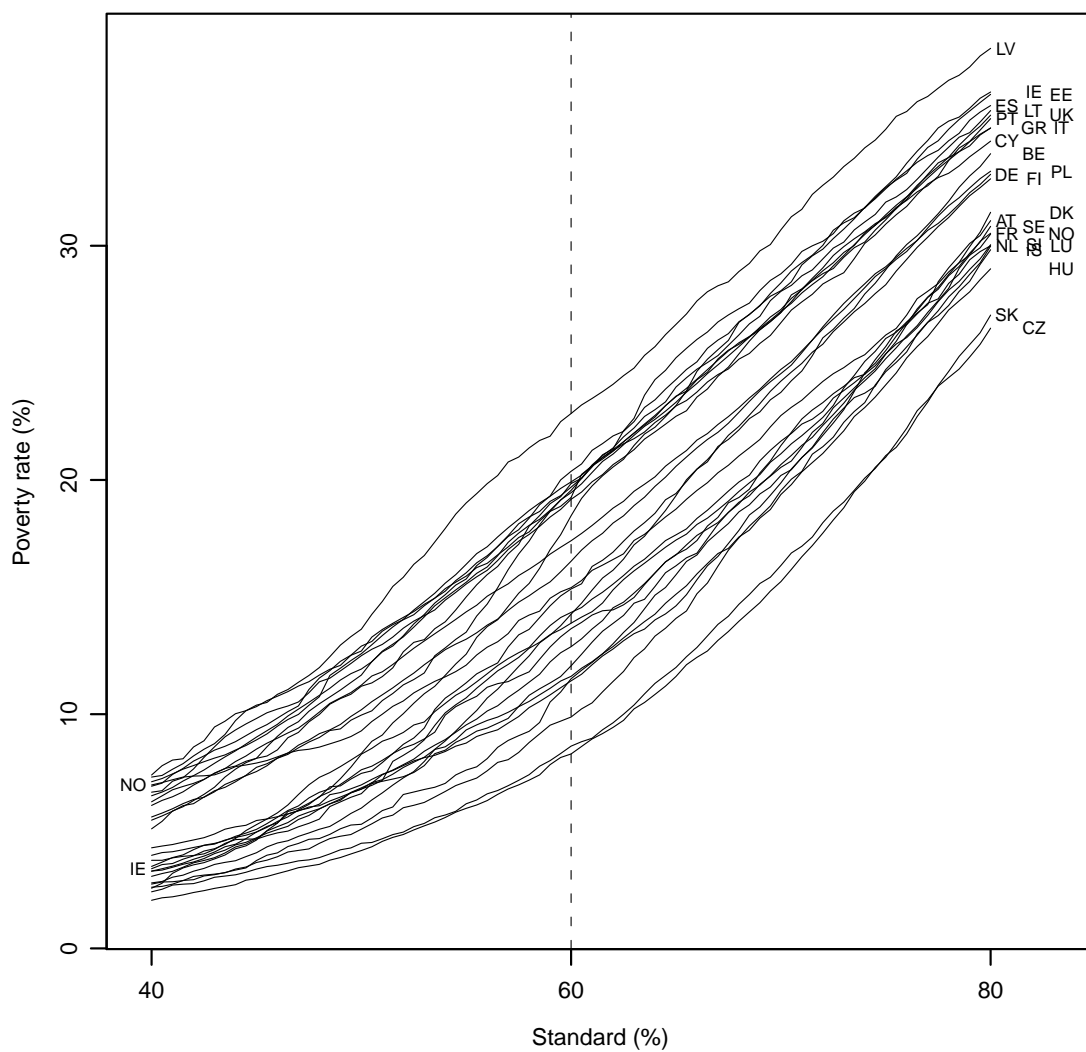
Figure 9: Sensitivity of the estimates of the household-level poverty rate with respect to the threshold of $40-80\%$ of the median equivalised household income.

The estimated household-level poverty rates are uniformly higher than the individual-level rates for 20 countries. Of the remaining six countries in EU-SILC, they are lower for Spain only in a short segment of the lowest threshold percentages ($40-42\%$), are lower for $50-60\%$ of the range of threshold percentages for Czech Republic, Slovakia and Hungary, and lower for 90% of the range for Luxembourg, and are lower throughout the range for Poland. In general, poverty is concentrated among the single-member and large households, but in most countries there are many more people living on their own, and the poor among them have a greater impact on the poverty rate.

3.1 Poverty rates in the regions

The sensitivity curves for the individual-level poverty rates in the regions of the four countries with respect to the thresholds of $40-80\%$ of the median household income are plotted in Figure 10. The regional poverty rates are evaluated with respect to a common national standard but, of course, the countries have different standards. The panels have identical vertical scales, so that comparisons could be made also across the countries. For example, the poverty rates in Spain are in a much wider range than in the other three countries. The sensitivity curves for one region in France (83, Corsica) and one in Spain (43, Extremadura) are curtailed, so as to maintain a good vertical resolution of the plots. For the standard of 80%, the estimated poverty rate in Corsica is 64.0% and in Extremadura is 57.3%. Some of the estimated curves are coarse, especially for France and Spain, because the corresponding regions have small sample sizes. The curve for Corsica is distinctly stepwise constant because the region is represented in the survey by only 29 households.

The panels show that the regional poverty rates are sensitive to the setting of the threshold. For example, in Czech Republic, Prague (region 1) has the lowest poverty rate, except for very low threshold, but the curves for regions 2, 3 and 6 intersect several times, as do the curves for regions 4 and 8.

For France, region 83 (Corsica) has the highest estimated poverty rate for every threshold, but the curves for the other regions intersect a great deal. In Spain, region 22 (Navarra) has the lowest poverty rate for every threshold and region 43 (Extremadura) the highest for thresholds between 55% and 80%. For the lowest thresholds, regions 63 (Ceuta) and 64 (Melilla) have the highest poverty rates, but the poverty rate of the
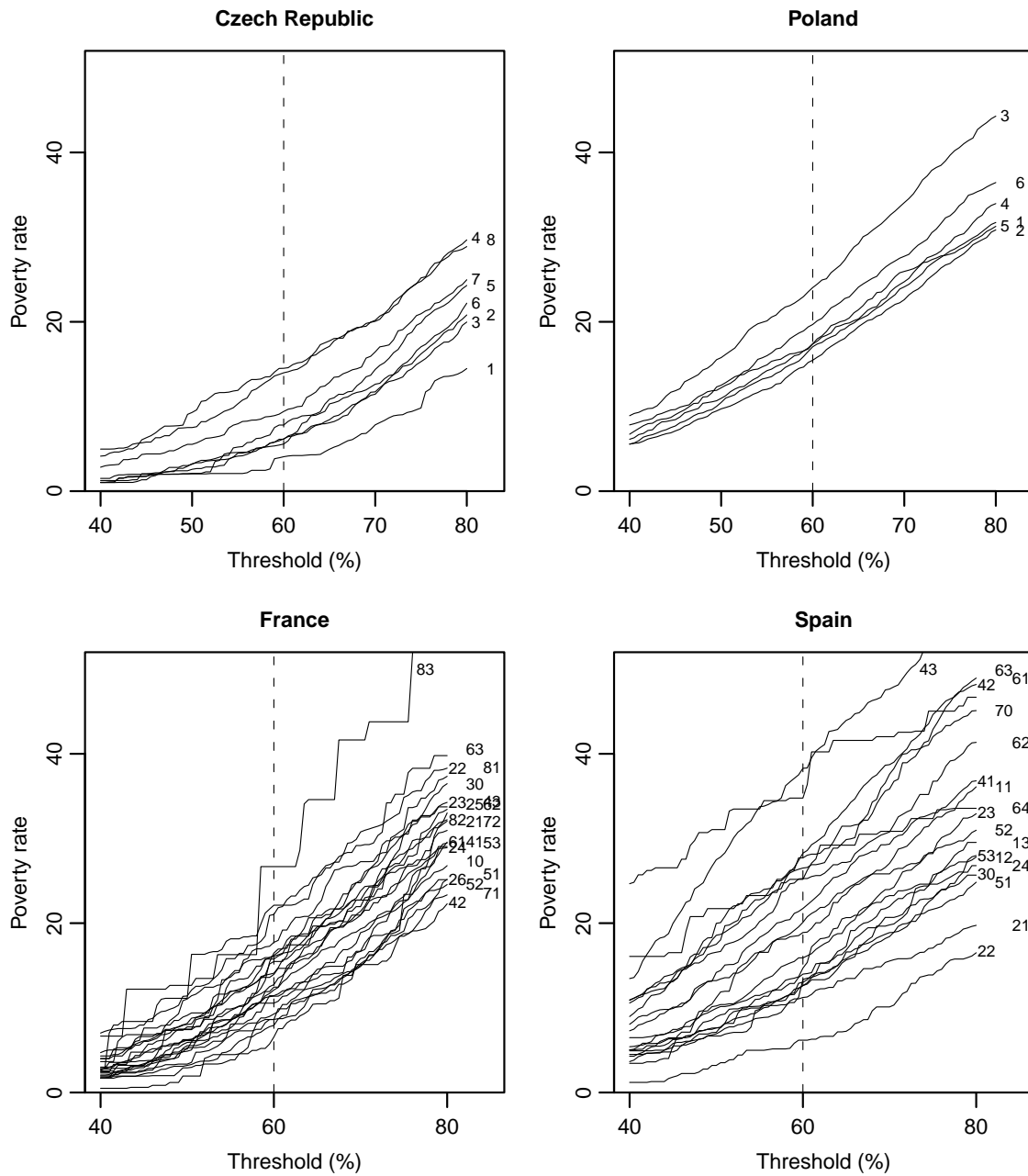
Figure 10: Sensitivity of the estimates of the individual-level regional poverty rates with respect to the thresholds of $40-80\%$ median equivalised household income.

latter in particular is close to the average for high thresholds. Ceuta and Melilla are two Spanish enclaves in northern Africa, bordering on Morocco. Their population and sample sizes are much smaller than for the other regions.

The sensitivity curves for the household-level regional poverty rates are plotted in Figure 11. They are very similar to the individual-level curves in Figure 10, but some differences can be discerned. For example, region 7 in Czech Republic (Central Moravia) has appreciably lower individual-level rates than regions 4 and 8 (Northwest Bohemia and Moravian Silesia) throughout the range of thresholds, but the household-level rates of these three regions are very similar throughout the range. The estimated household-level poverty rates in France are dispersed less than the corresponding individual-level rates, and the household-level rates for Corsica stand out more than the individual-level rates, even for the lowest threshold percentages.

4 Poverty in special subpopulations

Poverty tends to be more prevalent in households with a single parent and poverty is generally regarded as a more serious social ill when children are affected. In this section, we study the sensitivity of the poverty rates in these subpopulations with respect to the setting of the threshold. We reduce the range of thresholds that we explore to $50-80\%$, because otherwise we would observe extreme sensitivity.

The numbers of single-parent households, their percentages in the countries' samples, the numbers of individuals in these households, and the numbers of children up to 14 and 17 years of age are listed in Table 1. Single parents lead between 1.8% and 6.4% of the households. Children up to 14 years of age form between 11.4% and 21.5% of the countries' samples; the percentages of children up to 17 years of age are around $4-5\%$ higher. The counts and percentages of children up to 15 and 16 years of age, used in the analysis in Section 4.1, can be approximated quite reliably by interpolation. The sample sizes for these subpopulations are so small that an analysis for them in most of the regions would not be meaningful.

Figure 12 displays the sensitivity curves for the individual-level poverty rates of households led by single parents. The curves intersect and great deal — rates are highly sensitive to the threshold percentage. Denmark has the lowest estimated poverty rate
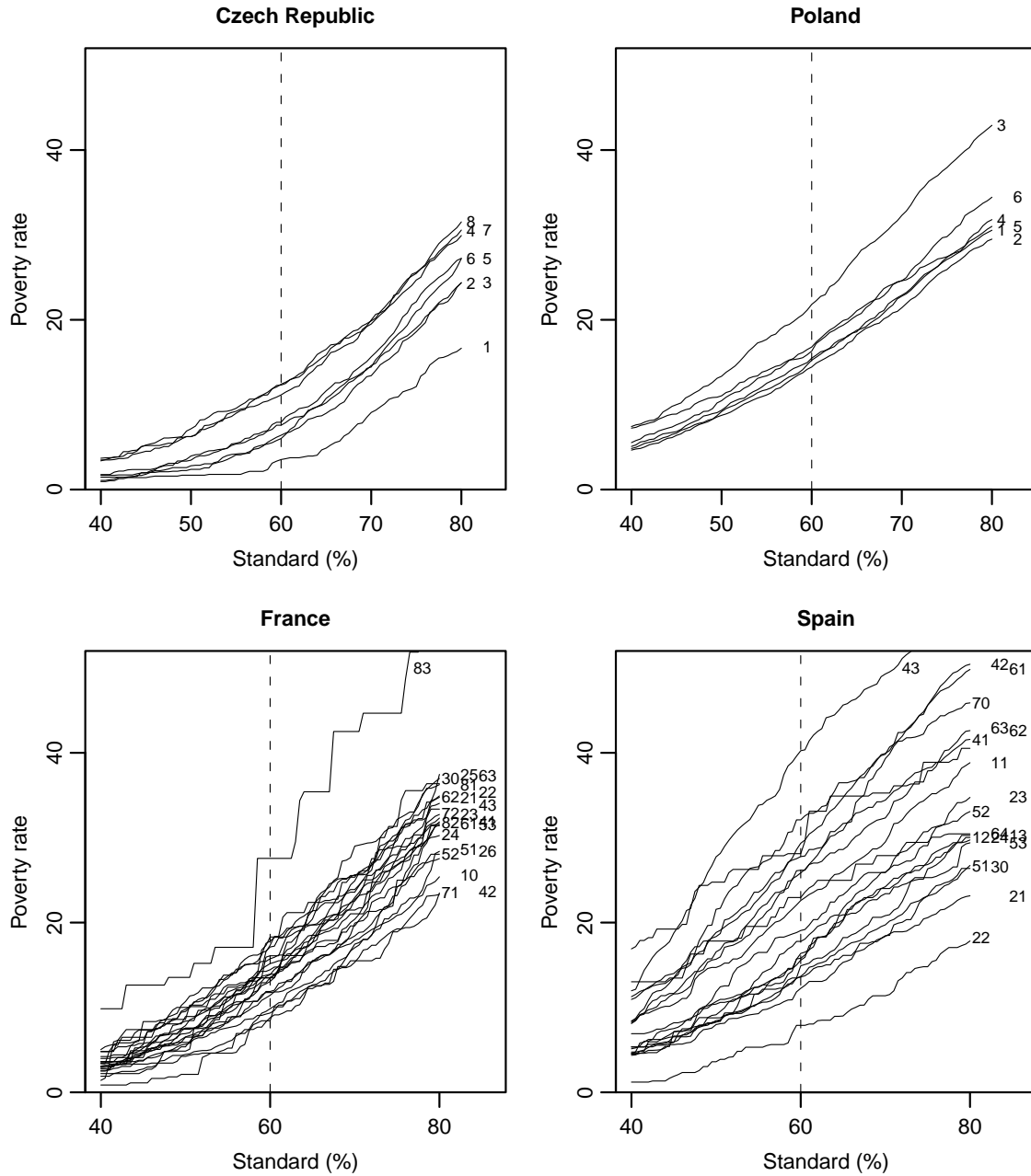
21

Figure 11: Sensitivity of the estimates of the household-level regional poverty rates with respect to the thresholds of $40-80\%$ median equivalised household income.

Table 1: The numbers of single-parent households and their members and the numbers of children and households with them. The percentages relate to the entire samples for the countries, of households for single parents, and of individuals for children.

|	Single parents			Children up to 14 years			Children up to 17 years		
	Households	%	Individuals	Children	%	In h-holds	Children	%	In h-holds
AT	352	5.17	883	2836	17.00	1749	3501	20.98	2055
BE	405	6.38	1101	2738	17.67	1584	3395	21.91	1906
CY	99	2.82	270	1830	17.22	1082	2324	21.86	1283
CZ	440	4.55	1113	3116	13.51	2088	3977	17.25	2498
DE	884	6.25	2129	4512	14.23	2877	5749	18.13	3508
DK	221	3.82	571	2820	18.94	1674	3575	24.01	2012
EE	294	5.71	748	1916	13.33	1299	2726	18.97	1679
ES	306	2.48	776	5013	14.49	3333	6174	17.85	3932
GR	104	1.84	266	2053	13.88	1264	2539	17.16	1527
IT	565	2.69	1407	7100	13.45	4779	8693	16.47	5619
FI	366	3.45	982	4701	17.12	2648	6255	22.78	3338
FR	589	5.61	1531	4825	18.62	2825	5937	22.92	3275
HU	380	4.35	961	3214	14.41	2056	4096	18.37	2467
IE	311	5.55	914	2485	18.15	1328	3100	22.64	1569
IS	152	5.30	397	1816	20.99	1111	2243	25.93	1303
LT	207	4.16	512	1486	11.63	1022	2075	16.24	1372
LU	203	5.23	554	2241	21.51	1345	2643	25.37	1518
LV	278	6.22	694	1415	12.62	993	1966	17.54	1292
NL	382	3.74	1005	5547	21.41	3059	6649	25.67	3536
NO	353	5.87	912	2954	19.52	1734	3761	24.85	2099
PL	520	3.64	1355	6623	15.46	4273	8640	20.16	5168
PT	99	2.30	255	1448	12.39	1035	1870	16.00	1276
SE	335	4.66	899	3283	18.11	1979	4354	24.02	2450
SI	241	2.77	614	3253	11.39	2155	4242	14.85	2699
SK	144	2.91	360	1706	11.48	1137	2367	15.93	1481
UK	537	5.79	1493	4013	18.29	2375	4927	22.45	2755
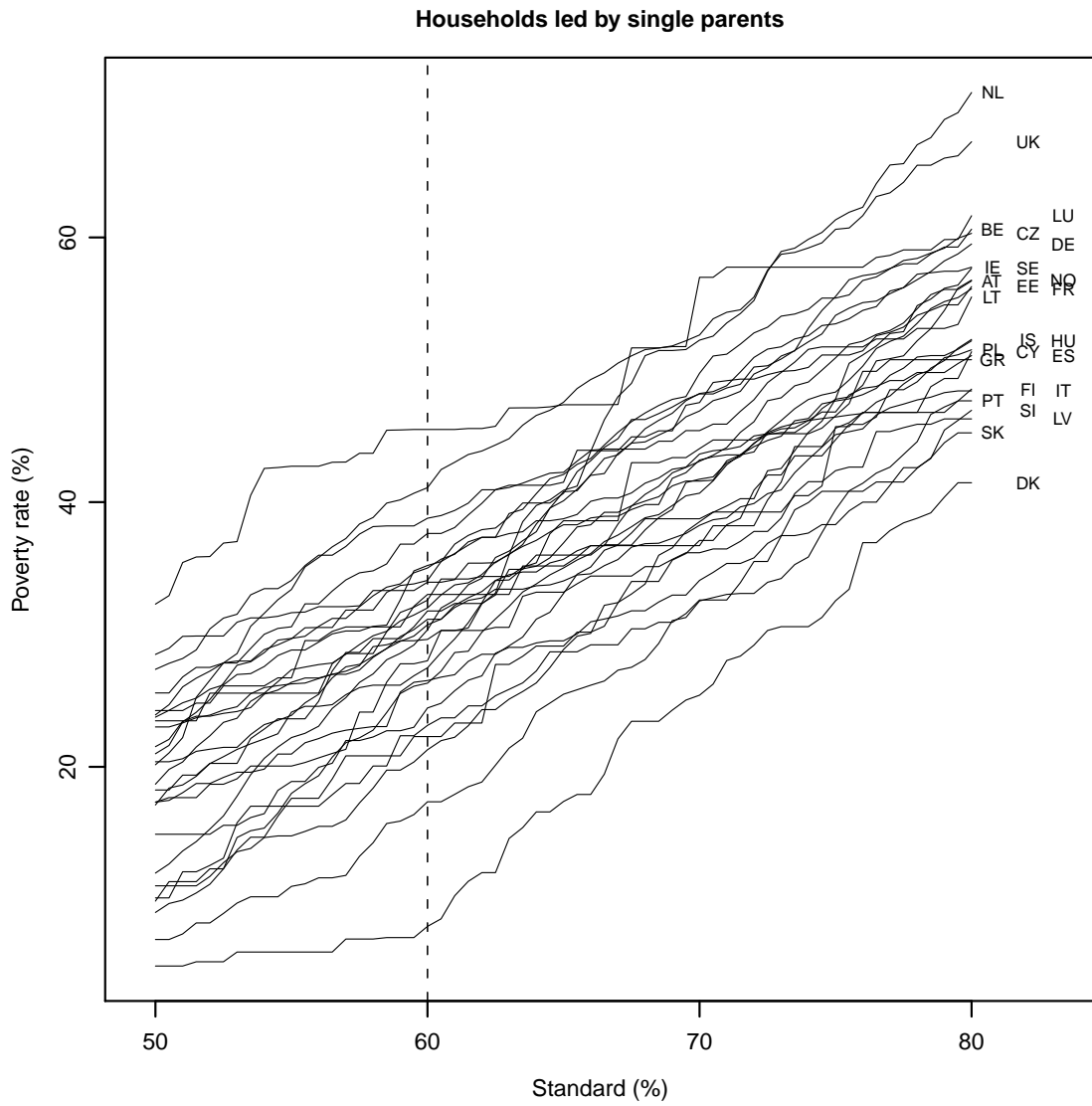
Figure 12: Sensitivity of the estimates of the national poverty rates for single-parent households.

for every threshold percentage, but some countries with high poverty rates at 80% threshold (the Netherlands and Belgium) have average or lower poverty rates at 50% threshold and some with low poverty rates at 80% threshold (Latvia and Portugal) have above average poverty rates at 50% threshold.

The extent of intersecting in Figure 12 can be summarised by the differences of the ranks of the national poverty rates at the highest and lowest threshold percentages. If the sensitivity curves did not intersect at all, these differences would vanish for all the countries. The largest possible difference, ±25, would be attained for a country that had the highest poverty rate at one extreme threshold percentage and the lowest at the other. The largest negative rank differences are attained for Portugal (–17), Spain (–16) and Latvia (–14) and the largest positive differences for the Netherlands and Belgium (16 each).

The household-level poverty rates are very similar to the individual-level poverty rates of households led by single parents; details are omitted.

4.1 Child poverty rates

We explore the poverty rates of children by the same approach, but add the age limit as another factor to the comparisons. It can be disputed whether the age of 15 or 16 is the appropriate upper limit for qualifying as a child, so we study the sensitivity with respect to the setting of the limit to $14-17$ years of age. The poverty rate for children is meaningfully defined only at the individual level.

The sensitivity curves of the national poverty rates of children with respect to the the thresholds $50-80\%$ are plotted in Figure 13 for the four age limits. The panels show that the rates are not sensitive to the age limit — the ranks of the countries are changed only slightly from one panel to another. In contrast, the poverty rates are quite sensitive to the setting of the threshold. The poverty rates rise steeply with the threshold percentage for Austria and France (from 8.1% to 37.6% and from 6.4% to 34.7%, respectively, for the age limit of 17 years), and at the slowest rate for Estonia, Latvia and Lithuania ($9.4-26.8\%$, $14.5-31.2\%$ and $11.4-29.4\%$, respectively). The rates are highest for all thresholds for Italy ($14.2-40.0\%$) and Poland ($17.1-39.1\%$) and lowest for Denmark ($5.2-20.6\%$). All the figures in parentheses are quoted for the limit of 17 years, but they differ from the corresponding figures for the other age limits
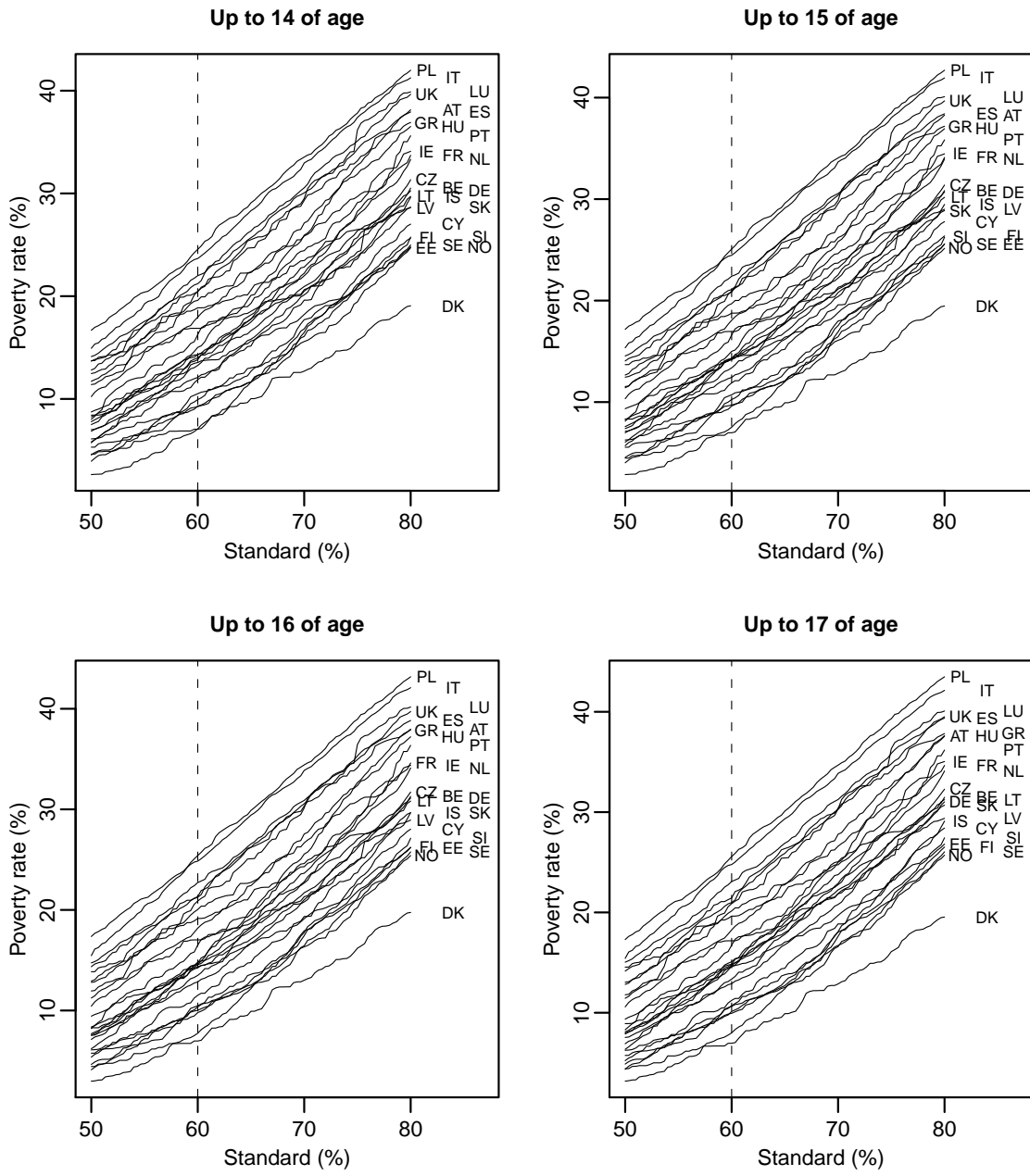
Figure 13: Sensitivity of the estimates of the national poverty rates of children with respect to the age limit and threshold.

by not more than 1%, and in many cases by much less. In conclusion, the poverty rates of children are highly sensitive to the setting of the threshold, even though the rates for some countries, when related to the others, do not change appreciably. The setting of the age limit is also unimportant.

Figure 14 presents the same summaries (curves of estimates), but with the four age-limit curves in the same panel for groups of countries. The age-limit curves are drawn by different line types, to confirm their insensitivity with respect to the limit. The countries are split into four panels both to avoid clutter, and to make their comparisons easier. Thus, Scandinavian countries have relatively low child poverty rates and they are in a narrow range. The poverty rates in the former Eastern-block countries are widely spread; they are among the highest (Poland) and the lowest (Slovenia for the threshold percentage up to 75% and Slovakia for higher percentage). There is a strange anomaly for Luxembourg, for which the estimated poverty rate steeply increases in a narrow range at the threshold of about 75%.

We emphasise that the effective sample sizes in these analyses are relatively small, because only single-parent households (all their members) and households with children (all members below the age limit) contribute to them, and so the sparseness of the data contributes to the appearance of high sensitivity. The sample sizes for these analyses are very small in many regions of France and Spain, so we do not discuss them.

5   Mean poverty gap

The poverty gap is defined as the shortfall of the (equivalised) household income with respect to the threshold. The poverty gap is equal to zero for households with income above the threshold. The sensitivity curves for the national mean poverty gaps are drawn in Figure 15. To improve the resolution of the diagram, the vertical axis is on the log scale. The mean poverty gaps are highly sensitive to the threshold; the curves rise most steeply (on the log scale) for Cyprus, Czech Republic, Slovenia and Luxembourg, and at the slowest rate for the Netherlands, Norway and Denmark.

Despite the high sensitivity, we can identify groups of countries with distinct ranges of mean poverty gaps. The former Eastern-block countries have the smallest mean gaps and, with a few exceptions, the most prosperous countries the largest mean gaps.
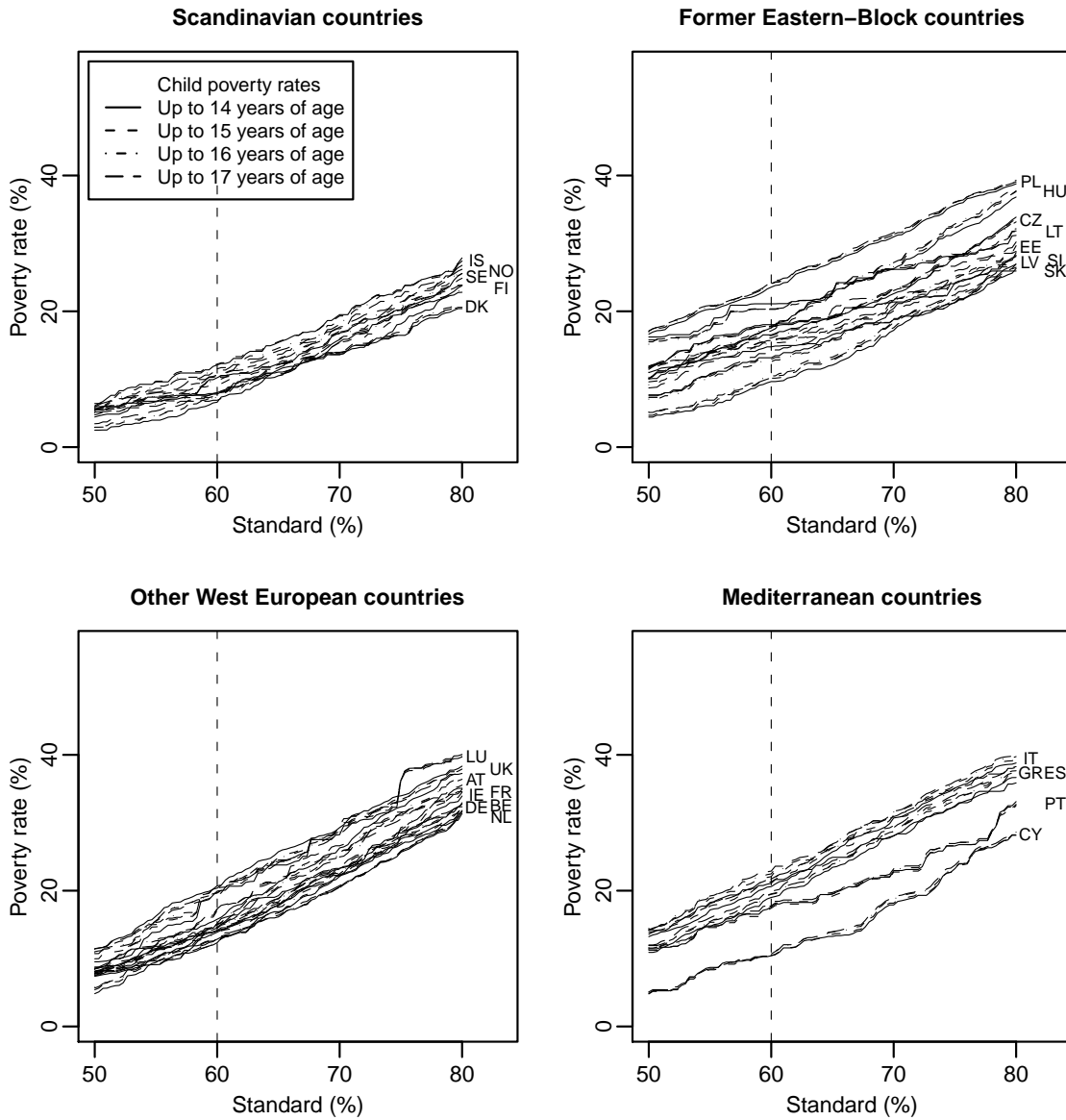
27

Figure 14: Sensitivity of the estimates of the national poverty rates of children with respect to the age limit and threshold, by groups of countries.

Figure 15: The national mean poverty gaps on the linear scale.

This pattern is an artefact of the (mean) levels of income in the countries, because a relatively small shortfall with respect to the poverty threshold amounts to a much greater 'poverty gap', as defined earlier, in a country with a high than in one with a low median household income. In brief, the pattern observed in Figure 15 is to a large extent a consequence of the misleading definition of the poverty gap.

We address this problem by an alternative definition which assesses the poverty gap on the log scale. Let E_i be the equivalised income of household i and T the threshold. Then the poverty gap (on the log scale) is defined as $-\log(E_i/T)$ if $E_i < T$ and zero otherwise. The mean poverty gap is estimated by the weighted average of the household-specific poverty gaps. In the averaging, we copy the household's poverty gap for every member of the household, and associate it with the individual's sampling weight. (The members of a household have identical weights). Of course, we have to discard households with negative income, for whom the poverty gap is not defined. We include all households with zero income by adding to the income one cent as a token.

The mean poverty gaps according to this definition, called the mean *log-poverty gaps*, are plotted in Figure 16 as functions of the threshold. They show a substantially different pattern and ranking than in Figure 15. For example, Lithuania and Latvia have high mean log-poverty gaps but low mean (linear) poverty gaps throughout the range of thresholds. Norway and Denmark have high linear poverty gaps but relatively much lower log-poverty gaps. The log-poverty gaps increase with the threshold percentage most steeply for Cyprus and Luxembourg, as in Figure 15, but also for Portugal, and they increase at the slowest pace for Sweden (unlike in Figure 15). Just as for the linear poverty gap, the log-poverty gap is also highly sensitive to the setting of the threshold.

The log-poverty ratios are on the scale of log-ratios (of two monetary amounts). Their means r are positive and small, in the range $0.00-0.12$. The negative exponentials $\exp(-r)$ are the (weighted) geometric means of the income of poor households divided by the threshold (and truncated from above by unity). Converted to the percentage scale, their complements $100\{1 - \exp(-r)\}$ are the mean relative shortfalls. Since the values of r are small, they can be approximated by $100r\,\%$, facilitating a simple interpretation of the mean log-poverty gap: $100r\%$ is the typical poverty gap (in the country or region) relative to the threshold for poverty.
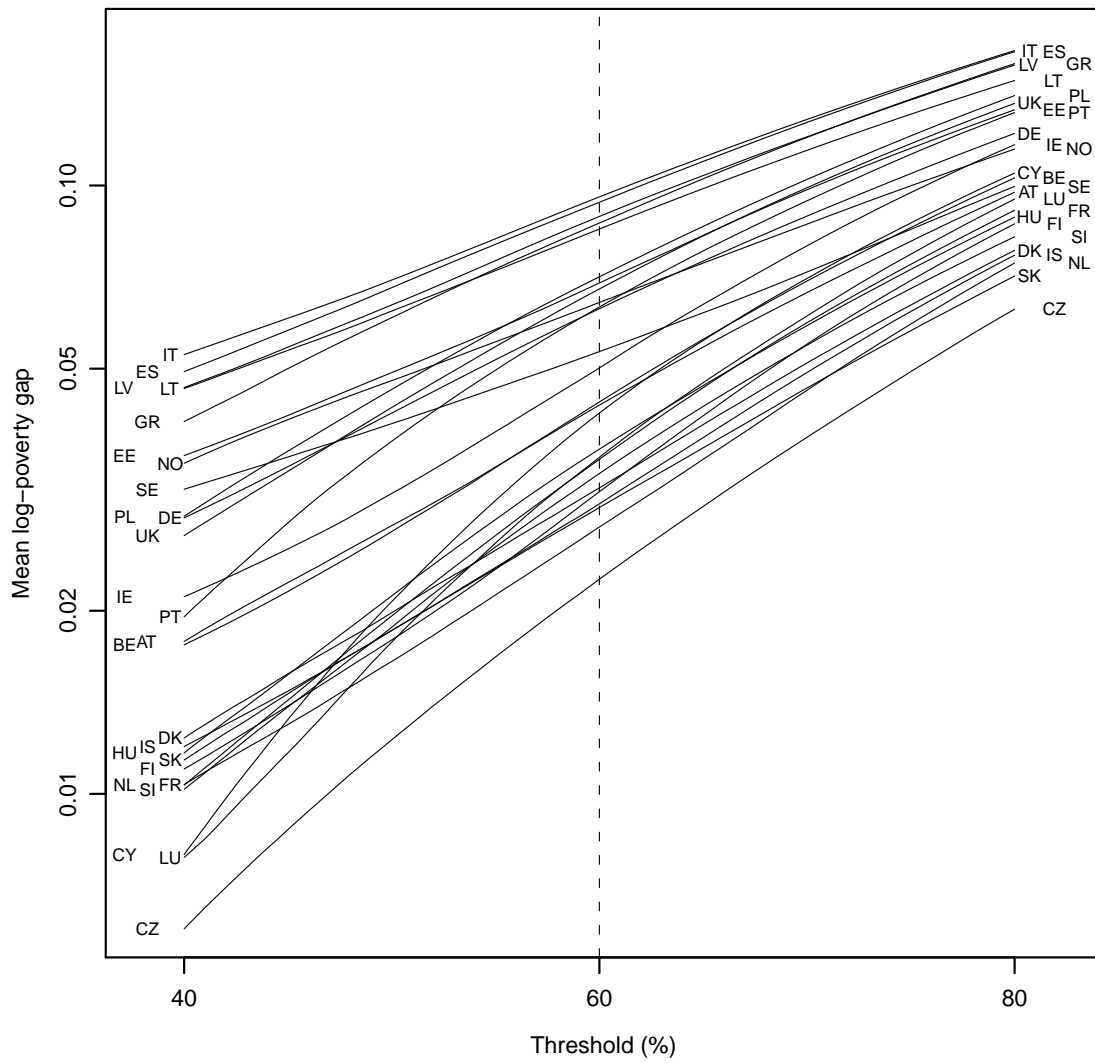
Figure 16: The national mean log-poverty gaps.

We do not argue that the log-poverty gap is more appropriate than the linear poverty gap, but merely want to point out that the scale on which the poverty gap is evaluated makes a lot of difference — the poverty gap is sensitive to the scale on which it is evaluated. The linear poverty gap is strongly associated with the median household income; a gap of several thousand Euro may be not serious in a country with high median household income, but it amounts to no income at all in a country with low median household income. We regard this as an obvious weakness of the linear poverty gap. The mean log-poverty gap is easy to translate to the percentage scale.

For completeness, Figure 17 displays the mean log-poverty gaps for the four countries with detailed information about their regions. The four panels have identical scales. To maintain a good resolution of the plots, the mean log-poverty gap for region 63 (Ceuta) is curtailed. Its value for the threshold of 80% is 0.43. The poverty gaps are very sensitive to the setting of the threshold. For Spain and France, we can identify groups of regions with distinct patterns of dependence of the poverty gap on the threshold. These differences are notable for Spain in the range $45\% - 55\%$ and for France in the range $40\% - 50\%$, where the poverty gaps intersect a great deal.

6   Trimming: from the mean to the median

Whether to estimate the mean or median, with or without weights, are often regarded as discrete choices that have to be made in the analysis of income in a population. In this section, we define a continuum of alternatives that have the mean and median as their extremes. We propose these alternatives not in order to select one of them, but to explore how much the estimates depend on the choice made.

As the alternatives, we propose the trimmed mean; see Figure 2. The (weighted) sample mean corresponds to no trimming and the (weighted) median to 50% trimming. Trimming $100p\%$ for $p \in (0, 0.5)$ at either extreme (tail) of the observed values can be motivated as a compromise between the mean and the median. In general, the sample mean is efficient when the studied variable is symmetric and does not have any outlying values. However, the sample mean is very sensitive to outliers. The sample median has complementary properties — it is resistant (not sensitive) to outliers, but
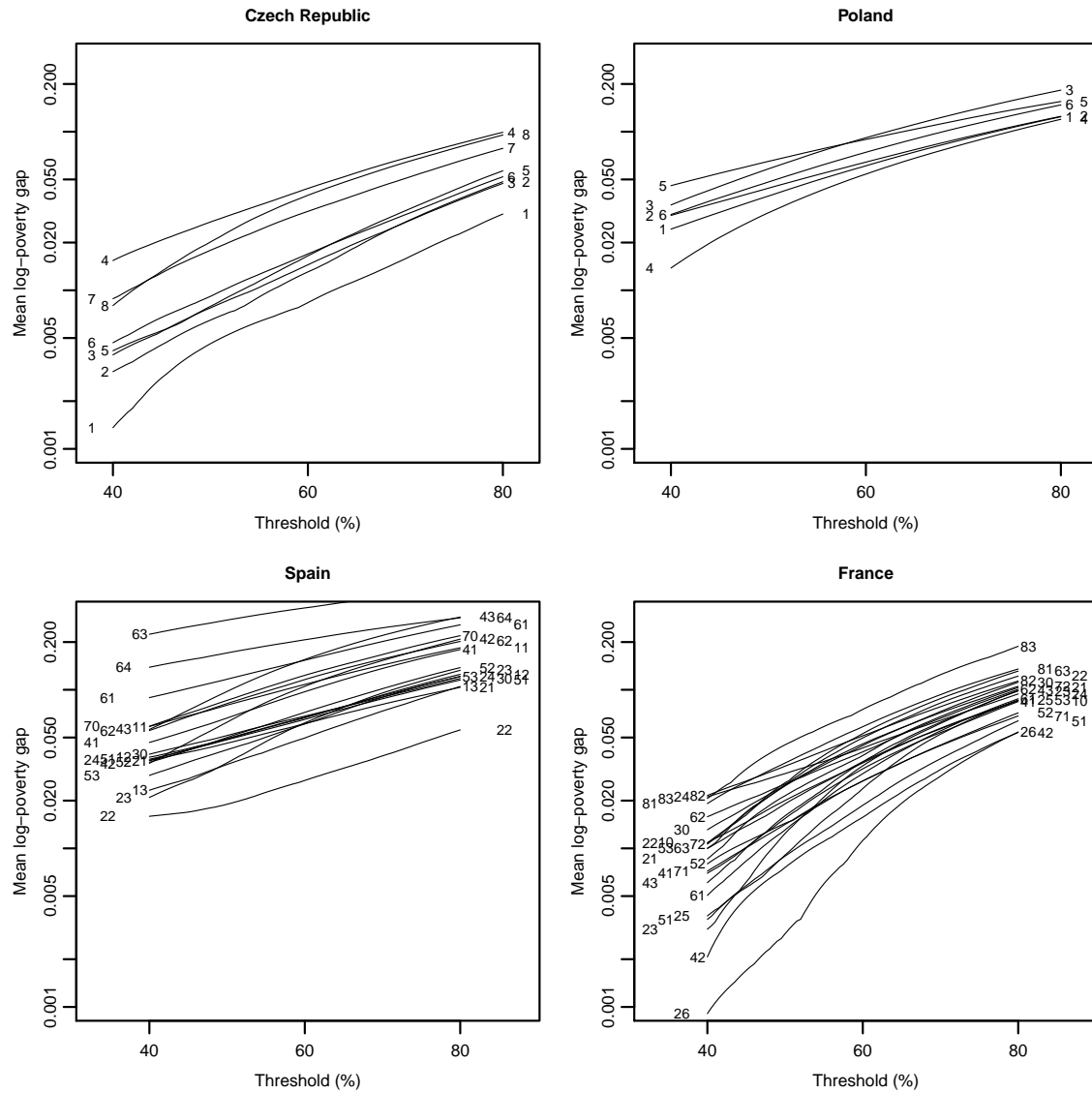
Figure 17: The regional mean log-poverty gaps.

is less efficient than the sample mean when the variable is close to symmetry and has no outliers.

We apply trimming, in the form of a sensitivity analysis, to the differences of equivalised household income in the panel component of EU-SILC for 2005 and 2006. Only a few countries have data from the two earlier years of the panel. In the analysis, we include only households that have records for both years and have no recorded changes in their membership between the two years. The households may have different equivalised sizes, because a member may become older than the age limit for children, and his or her contribution to the equivalised size then increases from 0.3 to 0.5.

We define the change of equivalised income as the difference of the equivalised income in the two years (2006 – 2005). The ratio of the two quantities may be considered, but difficulties would arise for households that have very small (or zero or negative) income in the 'denominator' year 2005. The estimates of the mean difference of equivalised income, as a function of the trimming percentage, are plotted in Figure 18. The estimated curves are presented in two panels, for the countries with high mean difference (over 1500 Euro) in the top and for the remainder in the bottom panel. This arrangement improves the resolution of the diagram. The only curve that straddles the two panels is for the Netherlands.

A small amount of trimming, say up to 5%, makes a lot of difference for most countries. The data for Belgium represent an extreme case, because one household had a very small income in one year and income of several million Euro in the other. The estimate without trimming is –516.60 Euro (off the scale in the bottom panel), but even for 0.5% it is 487.80 Euro, and it increases further with greater trimming until 7% (estimate 625.00 Euro), and then it decreases, first very gradually and then with a somewhat greater gradient. The median difference (50% trimming) for Belgium is 497.70 Euro. The estimated curve for the Netherlands has a very different shape; it decreases throughout the range of trimming, so that without trimming it would appear to have had one the highest increases in income, whereas with high trimming, or for the median, it has a relatively low difference (263.00 Euro). Even though the curves for the countries with the highest mean differences (top panel of Figure 18) are close to being parallel, except for very small trimming (up to about 3%), the estimates are highly sensitive to the extent of trimming.
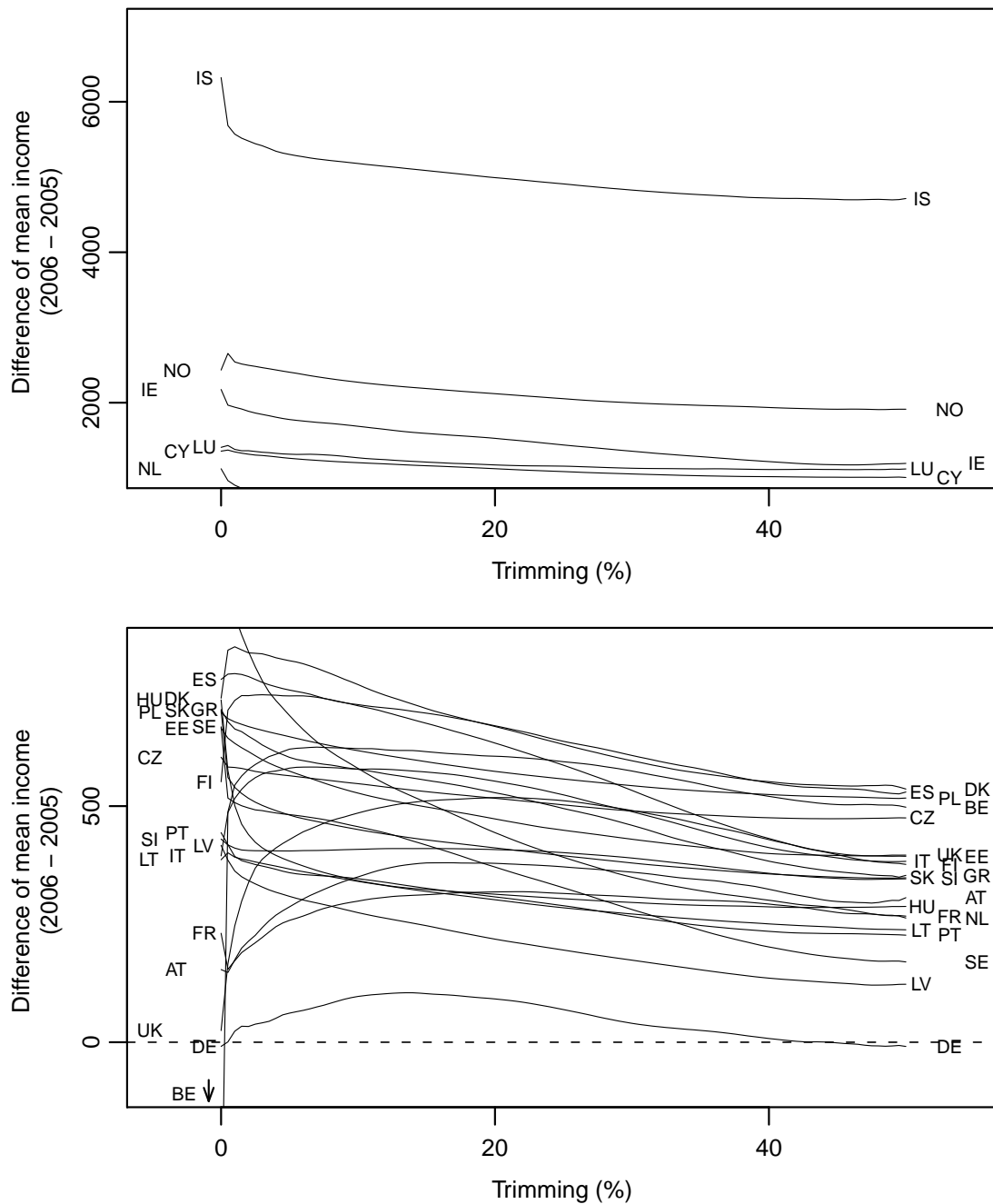
34

Figure 18: Sensitivity of the mean difference of income between 2006 and 2005 with respect to trimming.

The countries with high mean income tend to have high differences (although Cyprus, the UK, Austria and Germany are important exceptions), and so some form of scaling, such as dividing by the mean income may be useful. However, that would not alter our conclusion about the sensitivity of the estimates, because the shapes of the sensitivity curves would be changed only slightly.

The sensitivity of the estimates of the mean differences in income in the regions of the four countries with the available information is summarised in Figure 19. The sensitivity curves are curtailed at 5% because the estimates are extremely sensitive to trimming for small trimming percentage, and some of the panels would have to be drawn with very wide scales. Even after curtailing the plots at 5%, we cut off the vertical axis for France at −500 Euro, so that curves for two regions, 23 (Haute Normandie) and 83 (Corsica), are not displayed fully. Their estimates with 5% trimming are −1675.1 and −931.0 Euro, respectively, and the estimates are extremely sensitive for them even for much higher trimming percentages.

The estimates for the regions of Czech Republic and Poland are a bit less sensitive with respect to the trimming, but the choice of trimming, or between the mean and median, should be taken with care, because most estimation curves have wide ranges of values and are distictly not parallel. For example, region 1 (Prague) has by far the highest estimated mean increase in Czech Republic, but its estimated median differs only slightly from the estimated medians of three other regions. Similarly, region 1 in Poland (Lodz-Mazowieckie) has a relatively high estimated mean but nearly the lowest estimated median. Small sample sizes exacerbate the sensitivity of the estimates; that is why the estimates for the few regions of Czech Republic and Poland are less sensitive than the estimates for the regions of Spain and France. However, other factors that are much more difficult to pinpoint also contribute to sensitivity.

7   Transitions from and to poverty

Time spent in poverty is studied because long-term poverty has much more serious consequences than being poor intermittently and for only short periods separated by relative prosperity. The study of this aspect of poverty requires extensive data about income in periods of time shorter than one year. The available data permit us only
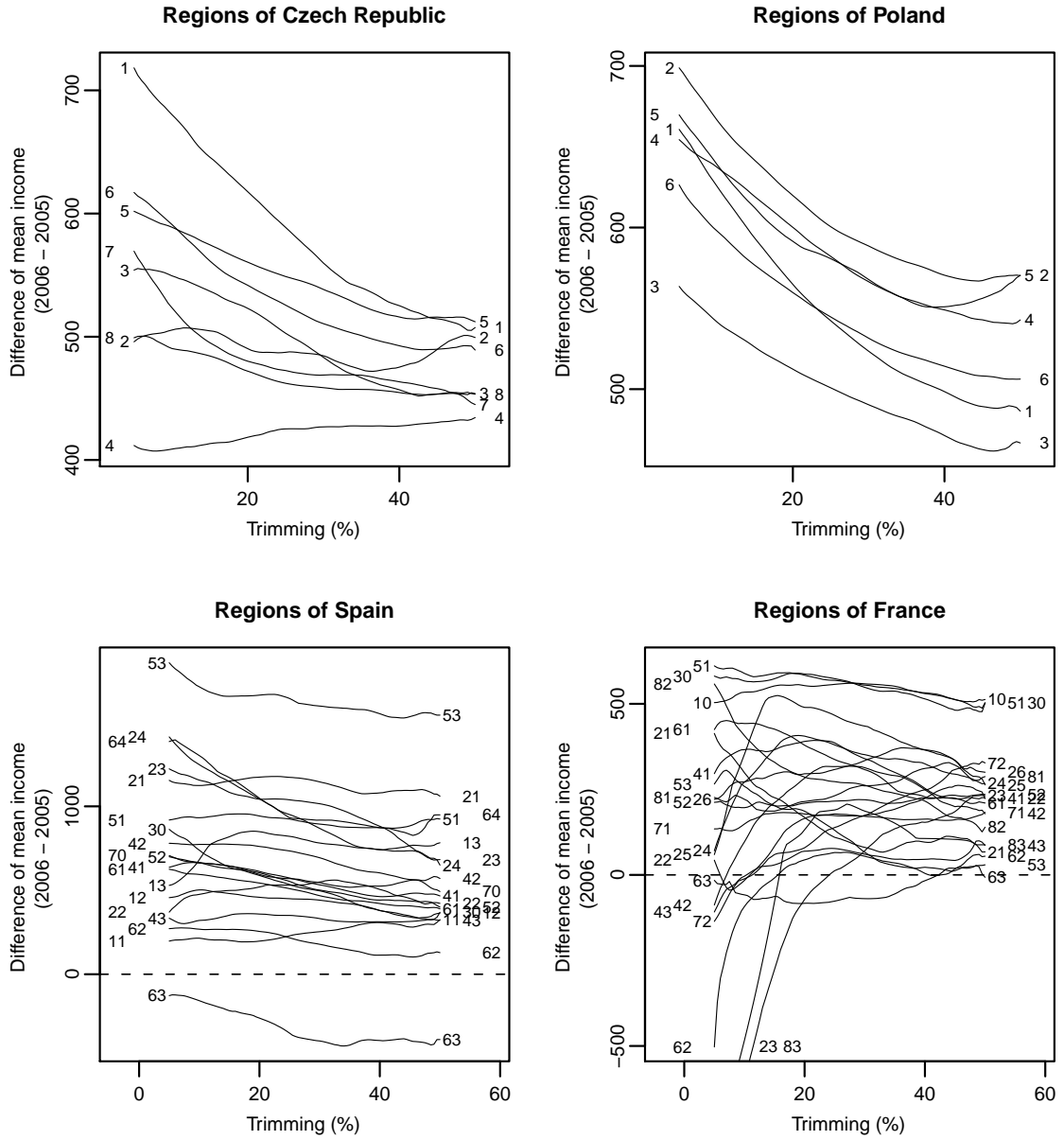
Figure 19: Sensitivity of the mean difference of income between 2006 and 2005 with respect to trimming in the regions of four countries.

to study the prevalence of switches between the poverty states from 2005 to 2006. Thus, we set the poverty thresholds for the two years and estimate the percentages of individuals (or households) who were classified as poor in 2005 and as not poor in 2006, and *vice versa*. These estimates are obtained for threshold percentages in the range 50–80%. The percentage of individuals who are classified as poor in both years is also of interest. For brevity, we refer to it as *persistence* of poverty. We reduce the sample to the households that have the same composition in the two years. If a member was classified as a child in one year and as an adult in the other, then the equivalised household size is altered, and this is reflected in the equivalised household income.

The transition rates are defined as the percentages of individuals who are in particular poverty states in the two studied years. The denominator for these percentages is the population size (or the total weight of the sample). An alternative definition is based on the conditional percentages, in which, for example, the transition rate from poverty (in 2005) to prosperity (in 2006) is defined as the percentage of individuals who made the transition among those who where classified as poor in 2005. To distinguish the two kinds of transition rates, we refer to them as unconditional and conditional.

The sensitivity curves for the unconditional transition rates between the poverty states and for being poor in both years 2005 and 2006 are plotted in Figure 20. In the left-hand panel, the estimated percentages of those who were classified as poor in 2005 and as not poor in 2006 are plotted in the top (positive) part of the panel, and the estimated percentage of those who made the transition in the other direction in the bottom (negative) part. A modicum of smoothing is applied to the estimated curves because the original estimates for some countries are quite erratic and have sudden changes for several threshold percentages. The curves tend to increase for both kinds of transition. However, even after smoothing, they are decreasing or have sudden bumps in several short intervals.

The curves intersect a great deal, so as indicators of change in the poverty status they are highly sensitive to the threshold percentage. The cause of the sensitivity is that estimation relies on small numbers of transitions and observations with large weights exercise unduly strong influence on the estimates. Despite the sensitivity, we can identify countries that have high prevalence of transitions in both directions
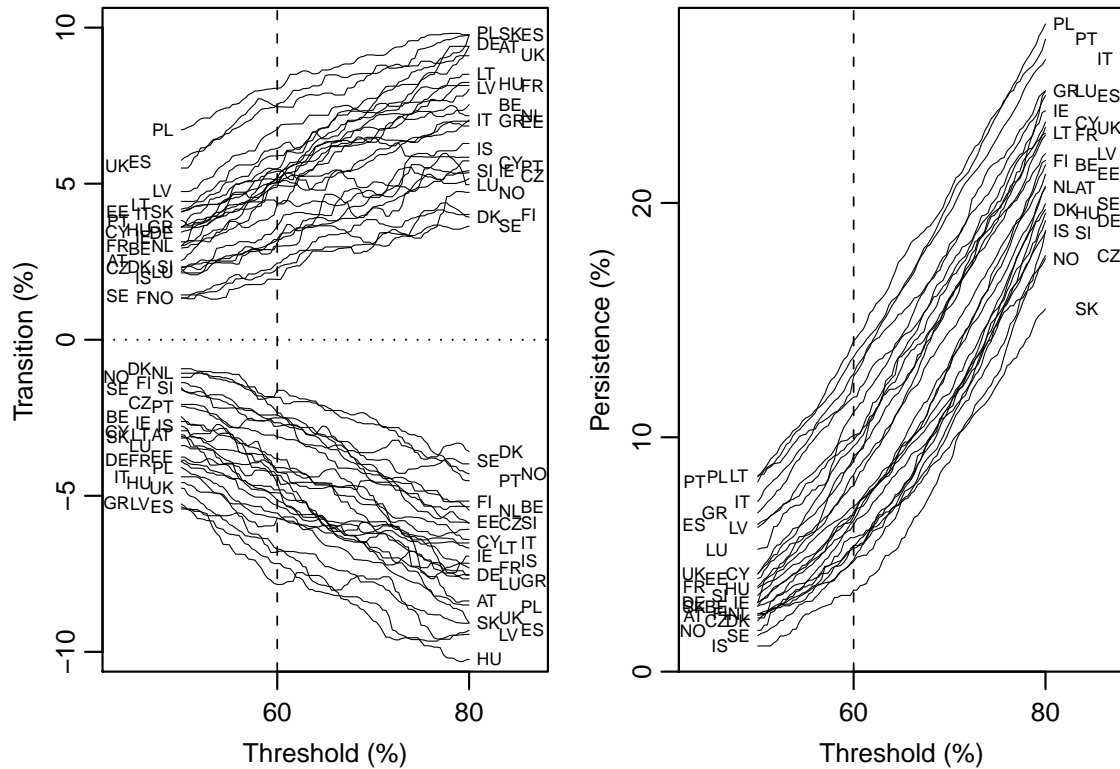
Figure 20: Sensitivity of the prevalence of transitions and of persistent poverty with respect to the threshold percentage.

(Hungary, Latvia, Poland, Slovakia and Spain), and those that have low prevalence (Denmark, Finland, Norway and Sweden).

The right-hand panel displays the estimated percentages of individuals who were classified as poor in both years. Of course, these percentages increase with the threshold. Sensitivity gradually decreases with the threshold percentage. For the threshold percentage below 50% they are very small and, being based on small effective sample sizes, are highly sensitive.

The sensitivity curves for the conditional transition rates are plotted in Figure 21. The left-hand panel is the 'conditional' version of the top part of the left hand-panel in Figure 20. The two sets of curves differ substantially, because the denominators for the conditional rates depend on the estimated poverty rates in 2005, and these vary a great deal. The unconditional rates tend to increase with the threshold percentage, whereas the conditional rates tend to decrease. The estimates of the conditional rates are more
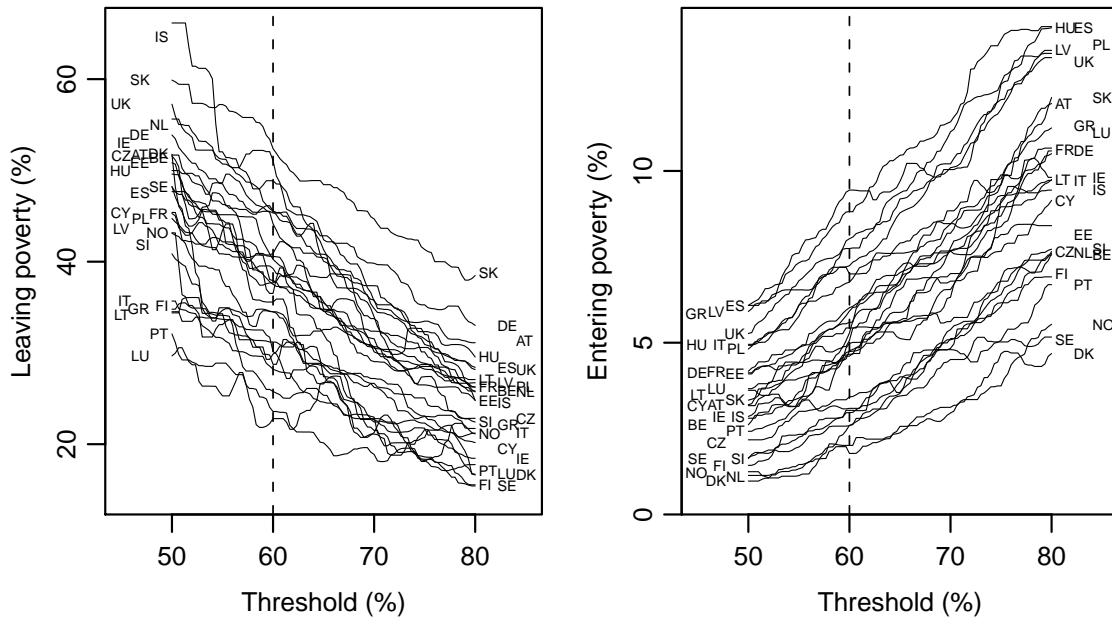
39

Figure 21: Sensitivity of the conditional transitions rates with respect to the threshold percentage.

unstable, have more and bigger deviations from the trend and are more sensitive to the threshold percentage. The countries' ranks of the conditional and unconditional transition rates for leaving poverty are very weakly related. For example, Iceland has the highest conditional rate of leaving poverty for threshold percentages up to 55%, but the corresponding unconditional rates are among the lowest. The unconditional rates for Norway are among the lowest, but the conditional rates are only slightly below the average.

The conditional and unconditional rates of transition to poverty are closely related and the ranks or relative positions of the countries in the corresponding diagrams are very similar. The reason for this similarity is that the rates of not being poor are much greater numbers than the rates of being poor, and so little difference arises between using those not being poor or the entire sample as the basis. In contrast, the two sets of rates for leaving poverty differ substantially because the poverty rates are relatively small numbers.

Similar conclusions are arrived at for the unconditional and conditional transition and persistence rates at the household level; details are omitted. For persistence rates,

40

it is not necessary to consider conditional rates because they are equal to the complement of the conditional transition rates for leaving poverty.

8   Discussion

Income and poverty status are basic variables in many studies of the economic well-being of a population. The definitions of such variables entail several conventions that, even though reasonable, do not have a profound scientific basis and have been set mainly for the sake of uniformity, comparability and convenience. We explored the impact of some of these conventions, such as the percentage of the median household income used as the poverty threshold and equivalisation of the household size, and identified several weaknesses in the definitions. If only slightly different conventions were adopted they would lead to substantially different conclusions (estimates) even about very simple population summaries, such as mean income and the poverty rate. This is to be expected, because poverty as a concept is on a continuum, and its definition as a dichotomy (Poor/Not poor) is adopted for mainly operational convenience. However, comparisons of estimates are altered substantially with small changes in these conventions.

Our perspective is not to revise the operational definitions, because there is no setting of the conventions that, by some credible objective criteria, is more valid than any other. Instead, we propose to accompany every substantive analysis with a sensitivity analysis which would repeat the former for several plausible changes of the conventional settings. If the results agree (or differ insubstantially) for all the settings, we have the common result as an unequivocal conclusion. Otherwise any single conclusion would be problematic and selecting one at the expense of the others amounts to a denial of uncertainty, or it has to be qualified by the selection made, thus limiting its relevance.

Sensitivity analysis is a generic term for a statistical exploration of the impact of settings and inputs on the conclusions. The settings are not necessarily captured or reflected in the data that are submitted to the analysis; in our case such an analysis entails generating 'new' (alternative) data for the outcome. By its nature, sensitivity analysis is never complete, because often there is a myriad of factors that could be explored. This calls for intelligent selection (prioritising) of the factors that should

41

be explored. Ignoring them all and regarding every operational definition or setting, however well established, as optimal is a poor solution.

Acknowledgements

References and bibliography

Atkinson, A. B., (1985). How should we measure poverty? ESRC Programme on Taxation, Incentives and the Distribution of Income. Discussion Paper 82.

Atkinson, A. B., (1987). On the measurement of poverty. *Econometrica* **44**, 749–764.

Atkinson, A. B., (1998). *Poverty in Europe.* Oxford, Blackwell Publishers.

Atkinson, A. B., Cantilon, B., Marlier, E., Nolan, B. (2002). *Social Indicators. The EU and Social Exclusion.* Oxford University Press, Oxford.

Central Statistical Office (2005). EU Survey on Income and Living Conditions (EU-SILC). First results, 2003, Statistical Release. Dublin, Ireland.

Delhausse B., Lüttgens, A., and Perelman, S. (1993). Comparing measures of poverty and relative deprivation. An example for Belgium, *Journal of Population Economics* **6**, 83–102.

Eurostat (2005). Material deprivation in the EU. Statistics in Focus. Eurostat, Luxembourg.

Eurostat (2009). EU-SILC User Database description (Version 2007.1 from 01–03–09). European Commission. Luxembourg.

Guio, A.-C. (2005). Income poverty and social exclusion in the EU25. Statistics in Focus. Population and Social Conditions. Eurostat, Luxembourg.

Lelkes, O., Zólyomy, E.,(2008). Poverty across Europe: The latest evidence using the EU-SILC Survey. European Centre.

Longford, N. T., and Pittau, M. P. (2006). Stability of household income in European countries in the 1990's. *Computational Statistics and Data Analysis* **51**, 1364–1383.

Lovillier, S., and Verger, D., (1997). Pauvreté d'existence, monétaire ou subjective sont distinctes. Economie et Statistique, INSEE, Paris.

Mejer, L., (2000). Social exclusion in the EU member states. Statistics in Focus Theme 3, 1/2000, Eurostat, Luxembourg.

Orshansky, M., (1963). Children of the poor. *Social Security Bulletin* **26**, 3–13.

Orshansky, M., (1965). Counting the poor: Another look at the poverty profile. *Social Security Bulletin* **28**, 3–29.

Pashardes, P., (2005). Poverty and social exclusion in Cyprus. OECD, Economic Policy Paper No. 07–03.

Pérez-Mayo, J. (2004). Consistent poverty dynamics in Spain. IRISS Working Papers No. 2004/09. Differdange, Luxembourg.

Pichaud, D., (2000). Poor relations. *The Guardian*, 1st December 2000. Manchester, UK.

Ravallion, M. (2003). The debate on globalization, poverty and inequality: Why measurement matters. World Bank Policy Research Working Paper No. 3038.

Ravallion, M., and Lokshin, M. (2003). On the utility consistency of poverty lines. World Bank Policy Research Working Paper No. 3157.

Sen, A., (1979). Issues in the measurement of poverty. *Scandinavian Journal of Economics* **81**, 285–307.

Sen, A., (1992). *Inequality Reexamined.* Oxford University Press. Oxford, UK.

Whelan, C. T., Nolan, B., and Maître, B. (2006). Measuring consistent poverty in Ireland with EU-SILC data. Working Paper No. 165, Economic and Social Research Institute, Dublin, Ireland.

Últims documents de treball publicats

NUM	TÍTOL	AUTOR	DATA
09.08	A sensitivity analysis of poverty definitions	Nicholas T. Longford Catia Nicodemo	Novembre 2009
09.07	Emissions distribution in postKyoto international negotiations: a policy perspective	Nicola Cantore, Emilio Padilla	Setembre 2009
09.06	Selection Bias and Unobservable Heterogeneity applied at the Wage Equation of European Married Women	Catia Nicodemo	Juliol 2009
09.05	La desigualdad en las intensidades energéticas y la composición de la producción. Un análisis para los países de la OCDE	Juan Antonio Duro Moreno Vicent Alcantara Escolano Emilio Padilla Rosa	Maig 2009
09.04	Measuring intergenerational earnings mobility in Spain: A selection-bias-free approach	María Cervini Pla	Maig 2009
09.03	The monetary policy rules and the inflation process in open emerging economies: evidence for 12 new EU members	Borek Vasicek	Maig 2009
09.02	Spanish Pension System: Population Aging and Immigration Policy	Javier Vázquez Grenno	Abril 2009
09.01	Sobre los subsistemas input-output en el análisis de emisiones contaminantes. Una aplicación a las emisiones de CH4 en Cataluña	Francisco M. Navarro Gálvez Vicent Alcántara Escolano	Març 2009
08.10	The monetary policy rules in EU-15: before and after the euro	Borek Vasicek	Desembre 2008
08.09	Agglomeration and inequality across space: What can we learn from the European experience?	Rosella Nicolini	Desembre 2008
08.08	Labor Supply Response to International Migration and Remittances in the Republic of Haiti	Evans Jadotte	Setembre 2008
08.07	Industrial districts, innovation and I-district effect: territory or industrial specialization?	Rafael Boix	Juny 2008
08.06	Why Catalonia will see its energy metabolism increase in the near future: an application of MuSIASEM	J. Ramos-Martin, S. Cañellas-Bolta	Juny 2008
08.05	Do creative industries cluster? Mapping Creative Local Production Systems in Italy and Spain	Luciana Lazzeretti, Rafael Boix, Francesco Capone	Març 2008
08.04	Los distritos industriales en la Europa Mediterránea: los mapas de Italia y España	Rafael Boix	Febrer 2008