CEPS INSTEAD

WORKING PAPERS

Ordinal Variables and the Measurement of Polarization

Alessio FUSCO[1]
Jacques SILBER[2]

CEPS/INSTEAD, Luxembourg[1]
Department of Economics, Bar-Ilan University, Israël[2]

Working Paper No 2011-33
May 2011

Ordinal Variables and the Measurement of Polarization*

Alessio Fusco
CEPS/INSTEAD, Luxembourg

Jacques Silber
Department of Economics, Bar-Ilan University, Israël

May 2011

Abstract

This paper aims at proposing measures of polarization for the distribution of a variable when information on the latter is only ordinal. The measures proposed are borrowed from the recent literature on the measurement of segregation. An empirical illustration is given, based on the *European Union Statistics on Income and Living Conditions* (EU-SILC) for the year 2008. The ordinal variable refers to the 'ability to make ends meet' and polarization is measured between groups defined by the citizenship of the household member who answered the household questionnaire. Results show that Luxembourg and Estonia have the highest degree of polarization whereas Cyprus, Ireland and the United Kingdom display the lowest degree.

Keywords: polarization; ordinal information; EU-SILC; segregation

JEL classification codes: D31; D63

* This paper was started when Jacques Silber was visiting professor at CEPS/INSTEAD. Core funding for CEPS/INSTEAD from the Ministry of Higher Education and Research of Luxembourg is gratefully acknowledged. The authors thank Gaston Yaloneztky for his very useful comments on a previous draft of this paper. Emails : alessio.fusco@ceps.lu; jsilber_2000@yahoo.com

1. Introduction

Much research has been devoted in recent years to the study of polarization, a concept which refers somehow to the clustering of incomes around local poles. The literature has generally made a distinction between two broad approaches to this topic. Polarization may on one hand take the form of bipolarization, a situation where there are many people who are very poor but also a significant class of very rich individuals. An important gap separates them so that in such a case there is no sizeable middle class. On the other hand polarization has also been linked to social conflict, the idea being that it is polarization rather than inequality which fuels social conflict. Those who view polarization as bipolarization and link it to the disappearance of the middle class have mainly extended the path breaking work of Foster and Wolfson (1992;2010), while those for whom the concept of polarization is related to the notions of social conflict, have based their analysis on the concepts of "identification" and "alienation" introduced in the very important contributions of Esteban and Ray (1994) and Duclos et al. (2004). Whereas the approach of Esteban and Ray (1994), as well as that of Zhang and Kanbur (2001), assumes that the groups are defined before measuring polarization, Duclos et al. (2004) extended the analysis by letting the data determine the groupings of individuals.

The novelty of the present paper is that it proposes measures of polarization in the case of an ordinal rather than cardinal variable. Let us suppose, for example, that a survey is conducted where individuals are asked to say whether they are able to make ends meet and the possible answers are (1) with great difficulty (2) with difficulty (3) with some difficulty (4) fairly easily (5) easily (6) very easily. Moreover it is assumed that information is also available on the population subgroup (e.g. ethnic group, gender, area of residence) to which the individuals belong. Borrowing ideas (e.g. Reardon, 2009) on the measurement of the inequality of the distribution of an ordinal variable between a given number of categories and assuming the presence of a certain number of (unordered) population subgroups, this paper considers that polarization is negatively related to the share of within groups inequality in the inequality of the distribution of the ordinal variable in the whole population (all subgroups being merged). Such an approach was in fact taken by Zhang and Kanbur (2001) who wrote that "...polarization measures discussed so far aim to capture the 'clustering' along the income dimension into high and low income groups. However, debates on polarization are often conducted in the framework of recognized and accepted non-income groupings. In the US, for example clustering of black and white income levels is as much concern as the disappearing middle class. In China…geographical clustering of income is a major policy concern…"

The paper is organized as follows. Section 2 reviews quickly the literature on the measurement of polarization in the case of a cardinal variable. Section 3 proposes then new measures of polarization in the case of an ordinal variable while Section 4 presents a short empirical illustration based on the European Union Statistics on Income and Living Conditions (EU-SILC) database for the year 2008.

2. On the measurement of polarization in the case of a cardinal variable

There are essentially two ways of apprehending the concept of polarization in the economic literature (see, Nissanov et al., 2010). A first approach puts the emphasis on the notion of bipolarization, a situation where you have a significant number of people who are very poor but you have also a significant amount of very rich individuals. There is a big income gap between these two groups and this probably implies that there is no sizeable middle class. Such an approach clearly defines the (two) groups on the basis of their income.

The second approach to polarization looks at the extent to which population is clustered around a small number of distant poles. The idea is that political or social conflict is more likely, the more homogenous, separate and of a similar size the groups are. Such a view was introduced by Esteban and Ray (1994) for whom society can be thought of as an amalgamation of groups, in the sense that two individuals drawn from the same group are assumed to be "similar" while two persons belonging to different groups will be considered as "different" with respect to a given set of attributes. Esteban and Ray (1994) postulate that polarization is related to two behavioral functions: identification and alienation. Identification is an increasing function of the number of individuals who are in the same income class as a given individual. In other words, an individual feels some degree of identification with those who are “close” to him. The alienation function on the contrary characterizes the antagonism caused by income differences: an individual feels alienated from those who are “far away” from him. While Esteban and Ray (1994) had assumed that the population subgroups were well defined, Duclos et al. (2004) let the data determine the various poles. In the present paper we will assume that the population subgroups are known, before analyzing the data. However, the subgroups will not be defined on the basis of the incomes of the individuals but on that of other socio-economic characteristics such as ethnicity.

Let us now take a closer look at the approach stressing the notion of bi-polarization. This concept was introduced in the economic literature by Foster and Wolfson (1992; 2010) and Wolfson (1994) who defined what they called polarization curves. Their first polarization curve turns out to be related to the concept of "increasing spread" while the second is also linked to the notion of "increased bipolarity". Without entering into details we can say that the idea of "increasing spread" implies that moving from the middle position (the median) to the tails of, say, an income distribution will make the distribution more polarized. This clearly implies that a rank preserving increment in incomes above the median or a rank preserving reduction in income below the median will widen the distribution, that is, increase the distance between the two groups (those above and below the median) and hence increase the degree of bi-polarization.

The concept of "increased bipolarity" refers on the contrary to the case where the incomes below the median or those above the median become closer to each other. This corresponds to a kind of "bunching" of the two groups in the sense that the gaps between the incomes below the median (or those above the median) have been reduced. In such a case bi-polarization is assumed to increase.

These two notions of "increased spread" and "increased bipolarity" explain why "inequality" and "bi-polarization" are two different notions. Whereas any regressive transfer (transfer from a poor to a richer individual) will increase inequality, it will increase the degree of bi-polarization if this transfer takes place across the median (that is, if money is transferred from an individual with an income below the median to one with an income above the median) but will decrease bi-polarization if it takes place on the same side of the median (if money is transferred from an individual with an income below the median to a richer individual whose income remains below the median income or if money is transferred from an individual with an income above the median to another richer individual).

Several cardinal measures have been proposed to evaluate the degree of bi-polarization (see, for example, Foster and Wolfson, 1992, Wang and Tsui, 2000, Rodriguez and Salas, 2003, Deutsch et al., 2007) as well as polarization (see, Esteban and Ray, 1994, Zhang and Kanbur, 2001, Duclos, Esteban and Ray, 2004, Lasso de la Vega and Urrutia, 2006, Esteban et al., 2007, Poggi and Silber, 2010).

Of particular interest, as far as the purpose of this paper is concerned, is the approach taken by Zhang and Kanbur (2001). They focus on two concepts: (i) the degree of homogeneity within each group; and, (ii) the degree of heterogeneity across groups. The idea is that high within-group homogeneity is bound to increase polarization while clear differences between two groups will also increase it. These two concepts can be quantified using the concepts of “within group inequality” (representing the spread of the distributions in the groups) and “between group inequality” (measuring the distance across the groups means) for decomposable inequality measures.[2]

[2] This conceptual framework can clearly be also re-interpreted in terms of the identification-alienation terminology mentioned previously: within group inequality represents a loss of identification and between groups inequality is a proxy for alienation.

Zhang and Kanbur (2001) proposed thus to use as polarization index the ratio of the between- to that of the within-groups inequality expressed as

$$ZK = (T_B / T_W) \qquad (1)$$

where T_B and T_W refer respectively to the between and within groups Theil indices. Since groups may be defined on the basis of a second characteristic (rather than on the basis of income only) the distributions of the different groups may overlap.

Deutsch et al. (2007), extending previous work of Berrebi and Silber (1989) on the measurement of the flatness of a distribution, defined a bi-polarization measure P_G as

$$\leftrightarrow P_G = (G_B - G_W) / I_G \qquad (2)$$

where G_B and G_W refer respectively to the between and within groups Gini indices while I_G is the value of the Gini index in the whole population.[3] Since in the case of non-overlapping groups the overall Gini index may be expressed (see, Silber, 1989) as

$$I_G = G_B + G_W \qquad (3)$$

we may combine (2) and (3) and derive that, in the case of non overlapping income groups, P_G may be written as

$$\leftrightarrow P_G = (G_B - G_W) / (G_B + G_W) \qquad (4)$$

Poggi and Silber (2010) extended this approach to the case of overlapping groups[4] and defined the index P_G in such a case as

$$\leftrightarrow P_G = (G_B - G_W) / (G_B + G_W + OVERLAP) \qquad (5)$$

where $OVERLAP$ refers to the residual of the decomposition of the Gini index by population subgroups (see, Silber, 1989).

---

[3] Deutsch et al. (2007) limited their analysis to the case of bi-polarization, that is, they assumed that the population is divided into two groups of equal size, the "poor" being those whose income is smaller than the median income and the "rich" those with an income higher than the median income

[4] It does not matter what the number of groups is.

Whether (in the case of several population subgroups) one adopts the formulation proposed by Zhang and Kanbur (2001) or that suggested by Poggi and Silber (2010), it is clear that polarization is assumed to increase with the between groups inequality and decrease with the within groups inequality. If we adopt a decomposable inequality measure, such as the Theil index, this implies that polarization could also be measured via the ratio of the between groups inequality to total inequality, that is, via the complement to one of the ratio of the within groups inequality to total inequality. This is precisely the approach that will be adopted in the next section where measures of polarization are proposed in the case of an ordinal variable.

3. On the measurement of polarization when the variable analyzed is ordinal

Let us assume, for example, that the variable analyzed is the one mentioned previously, namely one that is ordinal and which measures the standard of living of individuals on the basis of a question where they are asked to say whether they are able to make ends meet and given six possible (ordered) answers. For simplicity, we will still call "income" such a variable but in what follows it should be remembered that "income" refers to an ordinal and not a cardinal variable.

Suppose therefore that there is a $(I$ by $J)$ matrix whose lines i and columns j refer respectively to (unordered) population subgroups (e.g. ethnic groups) and to the (ordered) income category to which an individual belongs. The typical element X_{ij} of such a matrix will thus indicate the number of individuals who belong to ethnic group i and are in income category j. Call $X_{..}$ the sum $\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}$, $X_{i.}$ the sum $\sum_{j=1}^{J} X_{ij}$ and $X_{.j}$ the sum $\sum_{i=1}^{I} X_{ij}$ where I and J denote respectively the number of population groups and income categories. Call also $p_{i,k}$ the cumulative share $\sum_{j=1}^{k}(X_{ij} / X_{i.})$. We may therefore define the distribution (among the various income groups) of the individuals who belong to population subgroup i as $p_i = (p_{i,1},..., p_{i,j},..., p_{i,J})$

. Actually since $p_{i,J}$ is by definition equal to 1, we may assume that the income distribution of the individuals belonging to population subgroup i is well defined by the vector $p_i^{'} = (p_{i,1},..., p_{i,j},..., p_{i,J-1})$. The distribution of individuals in the whole population (including together all the population subgroups) will be defined by the vector $P^{'} = (P_1,..., P_j,..., P_{J-1})$ where $P_j = \sum_{i=1}^{I} w_i p_{i,j}$ and $w_i = (X_{i.} / X)$.

We have however to remember that we know only to which income category a given individual belongs so that we cannot really measure the inequality of the income distribution among individuals belonging to a given population subgroup, using traditional income inequality indices such as the Gini, Atkinson or generalized entropy indices. We could eventually transform the histogram describing such an income distribution, using the kernel density function technique, but this approach may be problematic for the first group and especially for the last group which generally does not have an upper bound.

There exists however an alternative solution. We know that these income categories are ordered so that we can use inequality indices which have been developed for the case of ordered categories (see, Abul Naga and Yalcin, 2008, Reardon, 2009, and Dutta and Foster, 2010). Allison and Foster (2004) were in fact the first to have stressed that using the traditional (cardinal or ordinal) approach to inequality measurement with ordered variables is problematic, to say the least. They recommended using a "median-based" ordering and argued that a distribution $\{z\}$ exhibits more inequality than a distribution $\{w\}$ if $\{w\}$ may be derived from $\{z\}$ via a sequence of "median preserving spreads". Allison and Foster (2004) limited their study to the analysis of the ranking of distributions of ordinal variables. Reardon (2009), on the contrary, proposed cardinal measures of the degree of inequality of the distribution of an ordered variable. Then, assuming that such a distribution may be observed for several population subgroups, he considered that segregation should measure the extent to which the degree of inequality of the distribution of an ordinal variable within unordered population subgroups is small when compared to the degree of inequality of the distribution of this variable in the whole population.

This idea of defining (between groups) segregation as the ratio of the between groups over the total inequality reminds us however of the way Zhang and Kanbur (2001) measured polarization, since they had assumed that the latter should be measured via the ratio of the between over the within groups inequality. This is why, we propose that the degree of polarization of the distribution of an ordinal variable should be related to the extent to which the ordered groups are evenly distributed across the unordered population subgroups (what Reardon had called 'evenness'). Borrowing mostly Reardon's (2009) ideas on the measurement of segregation as well as some of Silber and Yaloneztky's (2010) suggestions concerning the measurement of equality of opportunity when one of the variables is ordinal, we will assume that such a measure of polarization should have the following properties:

Maximum and minimum polarization of the distribution of an ordinal variable

- Between groups polarization will be minimal if within each population subgroup i the distribution of the individuals among the various ordered categories is equal to that in the whole population. In other words, in such a case, we should observe that $P_{i,j} = p_{i,j} \; \forall i, j$.
- Between groups polarization will be assumed to be maximal if within each population subgroup i all the individuals belong to the same ordered category so that $p_{i,j}$ is equal to either 0 or 1, for all population subgroups i and ordered categories j.[5]

Invariance of polarization to the total size of the population

[5] We evidently assume in this case that the category to which all the individuals in subgroup *i* belong, is not the same for the different groups.

If the number of individuals belonging to population subgroup i and to the ordered category j is multiplied by a constant μ, and if such a change occurs for all population subgroups and ordered categories, polarization should not vary. In other words the total size of the population does not matter. It is only the relative distribution of the ordered categories among the various unordered population subgroups which is important.

"Swap" of individuals between unordered population subgroups

Before explaining this assumption, we have to define some concept of dominance. We will say that the income distribution of population subgroup h dominates that of population subgroup k over the ordered categories f to l (where $1 \leq f \leq l \leq J$) if $p_{h,j} < p_{k,j}$ for all $j \in (f,...,l)$. In other words, population subgroup h dominates population subgroup k over the ordered categories f to l if there is a greater proportion of the population subgroup k than of population subgroup h at or below each ordered category from category f to l.

Given this definition of dominance of a population subgroup over another, we will now make the following assumption. If the distribution of population subgroup h dominates that for population subgroup k over the ordered categories f to l, and if an individual in the ordered category f could be moved from population subgroup k to population subgroup h while an individual belonging to the ordered category l would be moved from population subgroup h to population subgroup k, then polarization will decrease.

Table 1: Illustrating the concept of "swap" of individuals between unordered categories, assuming the ordered categories are income classes.

Case 1

Population Subgroup	**Lowest Income Class**	**Medium Income Class**	**High Income Class**	**Very High Income Class**	**Total**
A	14	16	26	44	100
B	38	28	22	12	100

Case 2

Population Subgroup	**Lowest Income Class**	**Medium Income Class**	**High Income Class**	**Very High Income Class**	**Total**
A	14	17	25	44	100
B	38	27	23	12	100

Case 3

Population Subgroup	**Lowest Income Category**	**Medium Income Category**	**High Income Category**	**Very High Income Category**	**Total**
A	15	16	25	44	100
B	37	28	23	12	100

The intuition here is that if we could (theoretically) "swap" individuals belonging to different income classes in a way that makes the distributions of the individuals (among the income classes) in two (unordered) population subgroups more similar to one another, then polarization should decrease. Table 1 gives an illustration. It is easy to observe that in Case 1, the distribution of the individuals belonging to population subgroup A dominates that of the individuals belonging to population subgroup B. If we could now move an individual belonging to population subgroup A from the high to the medium income class and another individual belonging to population subgroup B from the medium to the high income class (see case 2), we would conclude that polarization decreased because the disparity between the two population subgroups in the cumulative proportions of individuals at or below medium and high income classes has been reduced, while the other (cumulative) proportions did not change.

"Swap" of individuals between ordered categories

If the distribution of the individuals belonging to population subgroup h dominates that of the individuals belonging to population subgroup k over the income classes f to l, (where $1 \leq f \leq l \leq J$), and if an individual belonging to income class f is moved from population subgroup k to population subgroup h while an individual belonging to income class J is moved from population subgroup h to population subgroup k, then the resulting decrease in polarization will be greater than the one that would be observed if an individual belonging to income class f is moved from population subgroup k to population subgroup h while an individual belonging to income category l is moved from population subgroup h to population subgroup k.

The idea here is that the principle of "swap" should be sensitive to the ordering of the categories involved. "Swaps" of individuals who are farther apart (as far as their income class is concerned) should have a greater impact on polarization than "swaps" of individuals who are closer (as far as their income class is concerned).

This is illustrated in Table 1 by comparing Cases 1, 2 and 3. Case 3 is derived from Case 1 by moving an individual who belong to population subgroup A from the high income class to the low income class (and not to the medium income class as in Case 2) and another individual belonging to population subgroup B from the low income class (and not from the medium income class as in Case 2) to the high income class. As a consequence the decrease in polarization will be greater when we compare Cases 1 and 3 than Cases 1 and 2.

Merging Population Subgroups

If M unordered population subgroups are gathered into a smaller number of L unordered groups, then it will be assumed that polarization may be broken down into the sum of within and between population subgroups polarization.

Merging Ordered Income Categories

If F ordered categories are bunched into a smaller number of G broader ordered categories through the merging of adjacent categories, then it will be assumed that polarization may be broken down into the sum of polarization within and between the broader categories.

Measuring inequality in the case of ordered categories

As mentioned previously, the list of the assumptions that have been considered as desirable for measuring polarization in the case where only ordinal information is available concerning the variable under study was borrowed from work on the measurement of occupational segregation in the case where occupations may be ordered (see, Reardon, 2009). As mentioned previously, we combine here the idea that segregation could be measured as the ratio of between groups to total inequality (see, Reardon and Firebaugh, 2002, Watson, 2006, and Jargowsky and Kim, 2009) and the proposition of Zhang and Kanbur (2001) who recommended measuring polarization as the ratio of between over within groups inequality. In other words we will assume that polarization, in the case of ordered categories, amounts to measuring the ratio of the inequality between the unordered population subgroups to the overall inequality. Since we stated previously that overall inequality should be equal to the sum of the inequality between and within population subgroups, we may also state that polarization may be measured as the complement to 1 of the ratio of the inequality within unordered population subgroups to the overall inequality.

To measure the degree of inequality of the distribution of an ordinal variable within unordered population subgroups we will follow the work of

Reardon and Firebaugh (2002) on the measurement of multi-group segregation and make two assumptions. First we will consider that the degree of inequality of the distribution of an ordinal variable within unordered population groups is equal to the weighted sum of the inequality within the various unordered population groups, the weight of each group being equal to its share in the total population. Second we will follow again Reardon (2009) in defining our measure of the inequality τ_i of the distribution of an ordinal variable within a given unordered population subgroup. Before doing so we need to understand how inequality should be measured in the case of an ordinal variable. Reardon (2009) suggested that for an ordinal variable that includes J ordered categories (implying in our case that the income classes are labeled as $j = 1,..., J$) inequality will be assumed to be maximal (and thus normalized to 1) when half the population belongs to income category 1 and half to income category J. Conversely Reardon (2009) suggested that inequality will be minimal (and normalized to 0) when all individuals belong to some income class $j = j_0$ with j_0 varying from 1 to J.

As a consequence, as stressed by Reardon (2009), defining the degree of inequality of the distribution of an ordinal variable amounts to measuring how close the distribution of the ordinal variable will be to these minimal and maximal bounds.

As mentioned previously the income distribution of the individuals who belong to population subgroup i is well defined by the vector $p_i = (p_{i,1},..., p_{i,j},..., p_{i,J-1})$. At the light of what we wrote before we may therefore state that inequality will be maximal when the distribution of the individuals in the whole population will be defined by the vector $P = (P_1,..., P_j,..., P_{J-1}) = ((1/2),(1/2),...,(1/2))$. This corresponds evidently to the case where half the individuals belong to the poorest income class and half to the richest.

We may also note that there are J cases where there is no inequality at all, that is, where all the individuals belong to the same income class. In such a case we would write that $P = (0,0,...,0,1,1,1)$ and state that $P_j = 0$ for $j < l$ and $P_j = 1$ for $j \geq l$ where l can take any value between 1 and J.

Reardon (2009) recommends then to measure inequality as

$\tau = (\frac{1}{J-1})\sum_{j=1}^{J-1} f(P_j)$	(6)

where $f(P_j)$ is a continuous function defined on the interval [0,1] such that $f(P_j)$ is increasing when $P_j \in (0, \frac{1}{2})$, decreasing when $P_j \in (\frac{1}{2}, 1)$, maximal (with a value of 1) over the interval [0,1] when $P_j = \frac{1}{2}$ and minimal (with a value of 0) over the interval [0,1] when $P_j = 0$ or $P_j = 1$.

Reardon (2009) suggested four possible functional forms for f.

The first one[6] is written as

$f^1(P_j) = -[P_j \log_2 P_j + (1-P_j)\log_2(1-P_j)]$	(7)

The second one is expressed as

$f^2(P_j) = 4P_j(1-P_j)$	(8)

The third functional form is

$f^3(P_j) = 2\sqrt{P_j(1-P_j)}$	(9)

Finally the fourth functional form is

$f^4(P_j) = 1 - \lvert 2P_j - 1 \rvert$	(10)

We may then replace f in (6) with one of the functional forms given in expressions (7) to (10) and derive, as a consequence, four possible measures of variations which will henceforth be expressed respectively as τ_1, τ_2, τ_3 and τ_4.

[6] Note that in defining f^1 it is assumed that $(0\log_2 0) = \lim_{x \to 0}(x \log_2 x) = 0$

One may note that each of these measures of ordinal variation reaches its maximum value of 1 only in the case where half the population belongs to the poorest income category ($j=1$) and half to the richest income category $(j=J)$ so that the vector P will be expressed as $P=(\frac{1}{2},\frac{1}{2},...,\frac{1}{2},1)$. Similarly each of the four measures of ordinal variation will reach its minimal value of 0 only when all the individuals belong to the same income category. In such a case the vector P will be written as $P=(0,0,...,0,1,...,1)$.

Let us call s_i the share of unordered population subgroup i in the total population, with $s_i=(\sum_{j=1}^{J}X_{ij})/(\sum_{i=1}^{I}\sum_{j=1}^{J}X_{ij})$. We may then express polarization as

$$POLOR=\sum_{i=1}^{I}s_i\frac{\tau-\tau_i}{\tau} \tag{11}$$

where τ, which is defined in (6) refers to the degree of inequality of the distribution of the ordinal variable in the whole population while τ_i measures this inequality within population subgroup i. Note that expression (11) may be also written as

$$POLOR=1-\frac{\sum_{i=1}^{I}s_i\tau_i}{\tau} \tag{12}$$

where $\sum_{i=1}^{I}s_i\tau_i$ measures within population subgroups inequality while τ, as already stressed, measures inequality in the whole population.

We can now replace τ and τ_i in (12) with one of the functional forms defined in expressions (7) to (10) and hence derive four measures $(POLOR^1, POLOR^2, POLOR^3$ and $POLOR^4)$ of polarization, where

$$POLOR^1=1-\frac{\sum_{i=1}^{I}s_i\tau_i^1}{\tau^1} \tag{13}$$

with

$$\tau_i^1 = (\frac{1}{J-1})\sum_{j=1}^{J-1}\{-[p_{i,j} \log_2 p_{i,j} + (1-p_{i.j})\log_2(1-p_{i.j})]\} \tag{14}$$

and

$$\tau^1 = (\frac{1}{J-1})\sum_{j=1}^{J-1}\{-[P_j \log_2 P_j + (1-P_j)\log_2(1-P_j)]\} \tag{15}$$

Similarly

$$POLOR^2 = 1 - \frac{\sum_{i=1}^{I} s_i \tau_i^2}{\tau^2} \tag{16}$$

with

$$\tau_i^2 = (\frac{1}{J-1})\sum_{j=1}^{J-1}[4p_{i,j}(1-p_{i.j})] \tag{17}$$

and

$$\tau^2 = (\frac{1}{J-1})\sum_{j=1}^{J-1}[4P_j(1-P_j)] \tag{18}$$

For the third measure we write

$$POLOR^3 = 1 - \frac{\sum_{i=1}^{I} s_i \tau_i^3}{\tau^3} \tag{19}$$

with

$$\tau_i^3 = (\frac{1}{J-1})\sum_{j=1}^{J-1}[2\sqrt{p_{i,j}(1-p_{i.j})}] \tag{20}$$

and

$$\tau^3 = (\frac{1}{J-1})\sum_{j=1}^{J-1}[2\sqrt{P_j(1-P_j)}\ ] \tag{21}$$

Finally the fourth and last measure will be expressed as

$$POLOR^4 = 1 - \frac{\sum_{i=1}^{I} s_i \tau_i^4}{\tau^4} \qquad (22)$$

with

$$\tau_i^4 = \left(\frac{1}{J-1}\right)\sum_{j=1}^{J-1}\left[1 - \left|2p_{i,j} - 1\right|\right] \qquad (23)$$

and

$$\tau^4 = \left(\frac{1}{J-1}\right)\sum_{j=1}^{J-1}\left[1 - \left|2P_j - 1\right|\right] \qquad (24)$$

One may note the links between $POLOR^1$ and information theory (see, Theil, 1967), $POLOR^2$ and the concept of diversity (see, Lieberson, 1969) and $POLOR^4$ and the notion of dissimilarity (see, Duncan and Duncan, 1955).

Properties of the polarization indices

It is easy to verify that the four measures of polarization ($POLOR^1, POLOR^2, POLOR^3$ and $POLOR^4$) are bounded between 0 and 1. The maximum value of 1 is observed when for each population subgroup all the individuals belong only to one income class.[7] In such a case within population subgroups inequality is minimal and hence polarization, which is measured by the ratio of between to the overall variation, will be maximal.

One may also observe that polarization will be minimal when for each population subgroup the distribution of individuals among the various income classes is the same. This clearly implies that there is no polarization, since the distribution of the individuals among the income classes does not depend on the population subgroup to which they belong.

Note also that all four indices of polarization obey the principle of invariance to the size of the population since these indices will not vary when

[7] As already mentioned in footnote 5, it is evidently assumed that not all the population subgroups are concentrated in the same income class. Otherwise the classification into different income classes would be irrelevant.

the number of individuals in each cell of the matrix { X_{ij} } is multiplied by a constant.

One may also show (see, Reardon, 2009) that three of the four indices (what we have called $POLOR^1, POLOR^2$ and $POLOR^3$) obey the principle of "swap" of individuals between unordered categories, which was defined previously. One should not be surprised to find out that the index $POLOR^4$ does not obey this principle, since it is related to the dissimilarity index and it is well known in the literature on income inequality measurement, that the index of dissimilarity generally does not obey the principle of transfers.

Note that the indices $POLOR^1, POLOR^2$ and $POLOR^3$ obey the principle of "swap" of individuals between ordered categories and satisfy the additivity condition defined previously when mentioning the case of merging subpopulations. These polarization indices however do not obey the additivity condition defined when mentioning the case of merging ordered income categories (see, Reardon, 2009, for more details).

4. An Empirical Illustration

In this section, we use our proposed measures to assess the extent of polarization in terms of self-assessed ability to make ends meet between nationals and foreigners in European countries. We use the 2008 cross-sectional data from the European Union - Statistics on Income and Living Conditions (EU-SILC). EU-SILC is an instrument aiming at collective comparable cross-sectional and longitudinal micro-data on income and living conditions (Wolff et al., 2010). The micro-data collected at household and individual levels are meant to be representative of the population living in private households in each of the participating countries. EU-SILC is now the reference dataset for comparative analysis of income distribution and living conditions in the European Union (EU) and official statistics regarding the EU indicators of social inclusion (Atkinson et al, 2002) and the EU2020 social inclusion target (Atkinson & Marlier, 2010) are derived from this survey.

EU-SILC has a legal basis making its implementation in EU member State compulsory. The scope of EU-SILC is indeed defined by the Council and European parliament regulation 1177/2003 (amended by Regulations 1553/2005 and 1791/2006 to extend EU-SILC in new Member States) which provides all the necessary information about the definition of target variables, sampling rules, sampling sizes or tracing rules. However, it should be noted that EU-SILC is based on the idea of a common framework and is not a fully harmonized survey. The common framework consists of common procedures, concepts, rules and recommendations but flexibility is left in order for each country to integrate this new instrument into its own national system of social surveys.

EU-SILC was launched in 2003 in seven countries under a gentleman's agreement and was later gradually extended to 15 countries in 2004, 25 countries in 2005 and 2006, and to all the EU27 countries (plus Norway, Iceland and Switzerland) starting in 2008 on. Data on France, Germany, Malta, Slovenia and Switzerland were however not available in the EU-SILC Users' Database to which we had access. In addition, in our empirical analysis the unordered categories refer to the citizenship of the household member who answered the household questionnaire. We excluded countries in which there were less than 5% of non locals. Our working dataset is hence composed of 11 countries (see Table 2 for a list of the countries covered and their abbreviations).

The question we are interested in asks households' respondents whether they are able to make ends meet and six possible answers were proposed, as mentioned previously: (1) with great difficulty (2) with difficulty (3) with some difficulty (4) fairly easily (5) easily (6) very easily. This question appears in the household questionnaire so that we use the household as the unit of analysis (see Table A1 for the number of observations per country). Table 2 gives the distribution of this variable for each country as well as the median category.

Table 2: Ability to make ends meet (by country, in 2008)

Country	Nationality	great difficulty	difficulty	some difficulty	fairly easily	easily	very easily	Median category
AT (Austria)	All	5.1	9.6	26.9	27.8	22.5	8.2	4
	local	4.9	9.0	26.5	28.3	23.2	8.1	4
	other	6.6	16.9	31.5	21.4	14.2	9.5	3
BE (Belgium)	All	8.1	14.4	23.5	26.9	23.0	4.1	4
	local	7.7	14.0	23.4	27.4	23.5	4.0	4
	other	13.4	20.1	23.9	20.3	17.6	4.7	3
CY (Cyprus)	All	19.5	32.0	31.0	11.9	4.6	1.0	2
	local	19.6	32.1	31.6	11.9	4.0	0.8	2
	other	18.9	31.9	26.1	11.2	9.3	2.6	2
EE (Estonia)	All	3.4	9.3	28.9	50.4	7.4	0.6	4
	local	2.4	7.8	27.0	53.7	8.4	0.7	4
	other	8.5	17.1	39.3	32.9	2.1	0.0	3
ES (Spain)	All	12.2	17.0	30.6	26.8	12.4	1.0	3
	local	11.6	16.5	30.8	27.4	12.7	1.0	3
	other	23.3	27.1	27.1	16.0	6.2	0.3	2
GR (Greece)	All	20.8	33.8	26.0	13.3	5.3	0.8	2
	local	19.7	33.9	26.4	13.6	5.6	0.8	2
	other	37.8	32.6	19.9	8.6	0.9	0.2	2
IE (Ireland)	All	8.5	14.2	34.2	29.0	10.2	4.0	3
	local	8.2	14.2	33.7	29.5	10.3	4.1	3
	other	12.3	14.0	41.4	21.4	8.9	2.0	3
IT (Italy)	All	17.3	20.5	38.8	18.1	4.7	0.6	3
	local	16.9	19.9	39.2	18.6	4.8	0.6	3
	other	24.6	30.7	32.4	8.6	3.0	0.7	2
LU (Luxembourg)	All	2.0	4.8	12.6	29.7	40.1	10.8	5
	local	1.3	2.7	8.8	28.2	46.7	12.3	5
	other	3.0	8.1	18.9	31.9	29.6	8.5	4
LV (Latvia)	All	15.4	27.5	35.7	18.6	2.6	0.3	3
	local	13.8	26.0	36.7	20.3	2.9	0.4	3
	other	21.9	34.1	31.4	11.3	1.3	0	2
UK (United-Kingdom)	All	6.4	9.9	25.4	37.5	14.0	6.9	4
	local	6.4	9.8	25.3	37.5	14.0	6.9	4
	other	6.2	12.6	26.5	36.4	12.9	5.4	4

Source: EU-SILC 2008 cross-sectional data-files, version UDB 01.08.10 Authors' calculation.
The unit of analysis is the household

The country with the highest median is Luxembourg whereas Cyprus and Greece display the lowest median. There are clearly differences between the countries but it is important to recall that the proposed measures of polarization do not depend on the overall amount of ordinal variation in the population but on differences between unordered categories in the distribution of the variable under scrutiny. As already mentioned, the unordered categories refer here to the citizenship of the household member who answered the household questionnaire. There are two unordered categories: 'local' or 'other' (see Table A2 for the distribution of this variable by country). Table 2 gives also the distribution of the ability to make ends meets separately for the local citizen and the other household respondents. The medians for "locals" and "others" often differ within countries, being lower for foreigners ("others") than for locals in seven countries (AT, BE, EE, ES, IT, LU, LV).

Table 2 shows that in addition to variation across countries, there is also variation within countries. We now compute the indices of polarization defined in the previous section. The left panel of Table 3 presents the values of the four indices and the right panel present the normalized values of these indices so that the degree of polarisation of the country with the highest value of the index is 1 and that of the country with the lowest value is 0.

Note that the value for the index $POLOR^1$ is missing in Latvia (LV) because the distribution of the variable for the non locals in that country contains an empty cell so that a logarithmic formulation cannot be used.

Table 3: Value of the Polarization indices in 2008,

	Non normalized				**Normalized (value-min)/(max-min)**			
	$POLOR^1$	$POLOR^2$	$POLOR^3$	$POLOR^4$	$POLOR^1$	$POLOR^2$	$POLOR^3$	$POLOR^4$
AT	0.0031	0.0037	0.0022	0.0078	0.0955	0.1094	0.0707	0.0747
BE	0.0027	0.0031	0.0021	0.0097	0.0841	0.0905	0.0656	0.0929
CY	0.0030	0.0014	0.0041	0.0000	0.0924	0.0365	0.1353	0.0000
EE	0.0301	0.0313	0.0289	0.0716	1	0.9634	1	0.6859
ES	0.0065	0.0077	0.0050	0.0004	0.2109	0.2306	0.1680	0.0041
GR	0.0063	0.0065	0.0058	0.0000	0.2053	0.1942	0.1947	0.0000
IE	0.0018	0.0019	0.0015	0.0000	0.0519	0.0517	0.0465	0.0000
IT	0.0036	0.0044	0.0025	0.0066	0.1119	0.1296	0.0807	0.0635
LU	0.0298	0.0325	0.0239	0.1044	0.9892	1	0.8268	1
LV		0.0119	0.0118	0.0269		0.3615	0.4053	0.2579
UK	0.0002	0.0002	0.0002	0.0000	0	0	0	0
min	0.0001	0.0001	0.0001	0.0000				
max	0.0301	0.0325	0.0289	0.1044				

Source: EU-SILC 2008 cross-sectional data-files, version UDB 01.08.10. Authors' calculation.

To allow a better reading of the results, Table 4 presents the same results but here countries are ranked according to the value of each index. We highlighted in grey the countries mentioned before for whom the median for locals and for others differ. It appears that, whatever the polarization index, Luxembourg (LU) and Estonia (EE) are the two top ranked countries and the United Kingdom (UK) is the country with the lowest degree of polarization.

Table 4: Value of the Polarization indices in 2008,
with a separate classification of the countries for each index.

POLOR¹		***POLOR²***		***POLOR³***		***POLOR⁴***	
EE	0.0301	LU	0.0325	EE	0.0289	LU	0.1044
LU	0.0298	EE	0.0313	LU	0.0239	EE	0.0716
ES	0.0065	LV	0.0119	LV	0.0118	LV	0.0269
GR	0.0063	ES	0.0077	GR	0.0058	BE	0.0097
IT	0.0036	GR	0.0065	ES	0.0050	AT	0.0078
AT	0.0031	IT	0.0044	CY	0.0041	IT	0.0066
CY	0.0030	AT	0.0037	IT	0.0025	ES	0.0004
BE	0.0027	BE	0.0031	AT	0.0022	IE	0.0000
IE	0.0018	IE	0.0019	BE	0.0021	GR	0.0000
UK	0.0002	CY	0.0014	IE	0.0015	CY	0.0000
LV		UK	0.0002	UK	0.0002	UK	0.0000

Source: EU-SILC 2008 cross-sectional data-files, version UDB 01.08.10 Authors' calculation.

One may also note that the seven countries for which the median category is higher for the "locals" than the "others" (foreigners) are those where polarization is highest, when the index Polor4 is used.

In Table 5 we give the ranking of the different countries based on the data of Table 3; the country with the highest (resp. lowest) value of the polarization index receives a rank of 1 (resp. 10). We excluded Latvia (LV) for which no result was available for the *POLOR*^{1} index. For each country a Borda score is then obtained by adding up the rankings for each index of polarization. The Borda rank is then simply the ranking of the countries according to their Borda score (see, Qizilbash, 2004).

Table 5: Ranking of the different countries by polarization index in 2008,

	POLOR[1]	*POLOR*[2]	*POLOR*[3]	*POLOR*[4]	Borda score	Borda ranking
EE	1	2	1	2	6	1
LU	2	1	2	1	6	1
ES	3	3	4	6	16	3
GR	4	4	3	8	19	4
IT	5	5	6	5	21	5
AT	6	6	7	4	23	6
BE	8	7	8	3	26	7
CY	7	9	5	9	30	8
IE	9	8	9	7	33	9
UK	10	10	10	10	40	10

Source: EU-SILC 2008 cross-sectional data-files, version UDB 01.08.10. Authors' calculation. Countries are ranked according to their Borda score/ranking

These results confirm what had been previously observed. Luxembourg (LU) and Estonia (EE) are the two countries where polarization is highest and the United Kingdom (UK) is the country with the lowest degree of polarization. As could be expected, for three of the four countries (Cyprus (CY), Ireland (IE) and the United Kingdom (UK)) for which the median category is similar for the "locals" and for the “others" the Borda score/ranking is the highest (polarization is the lowest).

Finally, we assessed the rank robustness of our results computing multivariate concordance indices (see Seth and Yalonetzky, 2010). The Kendall's W coefficient of concordance (see, Kendall and Gibbons, 1990) is a non parametric statistic ranging from 0 (when there is no concordance in the ranking between the indexes) to 1 (when there is complete concordance. This index was found to be equal to 0.821 (with a boostrapped standard error of 0.112 for 1000 replications). We may therefore reject the hypothesis of independence between the countries and the polarization indices and conclude that there is a positive and significant correlation between the different polarization indices.

5. Conclusion

This paper aimed at proposing measures of the degree of polarization of the distribution of a variable when information on the latter is only ordinal. The measures proposed were borrowed from the recent literature on the measurement of segregation and derived from the idea that polarization is related to the existence of differences between the relevant unordered categories in the distribution of the ordinal variable analyzed.

An empirical illustration was provided which used data from the 2008 cross-sectional data of the European Union Statistics on Income and Living Conditions (EU-SILC). It appears that Luxembourg and Estonia have the highest degree of polarization when the ordinal variable under scrutiny refers to the ability to make ends meet and the (unordered) population subgroups to the ethnicity of the respondent. We also observed that the three countries (Cyprus (CY), Ireland (IE) and the United Kingdom (UK)) for which the median category is similar for the "locals" and for the “others" display the lowest degree of polarization, the UK having the lowest rank for the four indices. Finally we found that there was a positive and significant correlation between the different polarization indices.

Bibliography

Abul Naga, R. H. and T. Yalcin (2008), “Inequality Measurement for Ordered Response Health Data,” *Journal of Health Economics* 27(6): 1614-1625.

Allison, R. A. and J. Foster (2004) “Measuring Health Inequalities Using Qualitative Data,” *Journal of Health Economics* 23, 505-524.

Atkinson, T., Cantillon, B., Marlier, E. and B. Nolan (2002), *Social Indicators: The EU and Social Inclusion*, Oxford, Oxford University Press.

Atkinson A.B and E. Marlier (eds) (2010), *Income and Living Conditions in Europe*, Luxembourg, OPOCE.

Berrebi, Z. M. and J. Silber (1989) “Deprivation, the Gini Index of Inequality and the Flatness of an Income Distribution,” *Mathematical Social Sciences*, 18: 229-237.

Deutsch, J., M. Hanoka and J. Silber (2007) "On the Link between the Concepts of Kurtosis and Bipolarization," *Economics Bulletin* 4(36): 1-6.

Duclos, J.-Y., J. Esteban and D. Ray (2004) "Polarization: Concepts, Measurement, Estimation," *Econometrica* 72: 1737-1772.

Duncan, O. B. and Duncan, B. (1955) "A methodological analysis of segregation indexes," *American Sociological Review* 20: 210-217

Dutta, I. and J. Foster (2010) " Inequality of Happiness in US: 1972-2008," mimeo, 2010.

Esteban, J.-M. and D. Ray (1994) “On the Measurement of Polarization,” *Econometrica*, 62: 819-852.

Esteban, J.M., Gradin, C., and Ray, D. (2007) "An Extension of a Measure of Polarization, with an application to the income distribution of five OECD countries," *Journal of Economic Inequality* 5(1): 1-19.

Foster, J. E. and M. C.Wolfson (1992) "Polarization and the Decline of the Middle Class: Canada and the U.S.," mimeo.

Foster, J. E. and M. C.Wolfson (2010) "Polarization and the Decline of the Middle Class: Canada and the U.S.," *Journal of Economic Inequality* 8: 247-273.

Jargowsky and Kim (2009), “The information theory of segregation: uniting segregation and inequality in a common framework,” in Y. Flückiger, S. F.

Reardon and J. Silber, editors, *Occupational and Residential Segregation*, vol.17 of *Research on Economic Inequality*, Emerald, Bingley, UK.

Kendall, M. and J. D. Gibbons (1990) *Rank correlation methods* (Fifth ed.). Edward Arnold.

Lasso de la Vega, M.C. and A. M. Urrutia (2006) "An Alternative Formulation of the Esteban-Gradín-Ray Extended Measure of Polarization," *Journal of Income Distribution* 15 (Index Issue): 42-54.

Lieberson, S. (1969) ''Measuring Population Diversity,'' *American Sociological Review* 34:850–62.

Nissanov Z., Poggi A. and J. Silber (2010), "Measuring bi-polarization and polarization a survey," in Deutsch J. and Silber J. (eds). *The Measurement of Individual Well-Being and group Inequalities. Essays in memory of Z.M. Berrebi*, chapter 3.

Poggi A. and J. Silber (2010), "On polarization and mobility: a look at polarization in the wage-career profile in Italy," *Review of Income and Wealth*, Series 56(1):.123-140.

Qizilbash M. (2004), "On the arbitrariness and robustness of multi-dimensional poverty rankings," *Journal of Human Development*, 5(3): 355-375.

Reardon, S. F. (2009) "Measures of Ordinal Segregation," in Y. Flückiger, S. F. Reardon and J. Silber, editors, *Occupational and Residential Segregation*, vol. 17 of *Research on Economic Inequality*, Emerald, Bingley, UK, pp. 129-55.

Reardon, S. F. and G. Firebaugh (2002) "Measures of multi-group segregation," *Sociological Methodology* 32: 33-67.

Rodriguez, J.G. and R. Salas (2003) "Extended Bi-polarization and Inequality Measures," *Research on Economic Inequality* 9: 69-83.

Seth and Yalonetzky (2010) "Assessing Rank Robustness with Multivariate Concordance Indices: An Application to Multidimensional Poverty in India," mimeo.

Silber, J. (1989) "Factor components, population subgroups and computation of the Gini index of inequality," *Review of Economics and Statistics* 71: 107-115.

Silber, J. and G. Yalonetzky (2010) "On Measuring Inequality in Life Chances when a Variable is Ordinal", mimeo.

Theil, H. (1967) *Economics and information theory.* Amsterdam: North-Holland.

Wang Y.Q. and K.Y. Tsui (2000), 'Polarization Orderings and New Classes of

Polarization Indices,' *Journal of Public Economic Theory* 2: 349-363.

Watson T. (2006), *Metropolitan growth, inequality, and neighborhood segregation by income*, Williamstown, MA, Williams College.

Wolff P., Montaigne F. And G. Rojas Gonzélez (2010), "Investing in statistics: EU-SILC", in Atkinson A.B and E. Marlier (eds), *Income and Living Conditions in Europe*, Chapter 2, pp. 37-56. Luxembourg, OPOCE.

Wolfson, M.C. (1994) : 'When Inequalities Diverge' , *American Economic Review , Papers and Proceedings* , vol. 84 , pp. 353-358.

Zhang, X. and R. Kanbur (2001), 'What Differences Do Polarization Measures Make? An Application to China' , *Journal of Development Studies*, 37(3):85-98.

Table A1. Number of observations by country

Country	Frequency
AT	5,707
BE	6,279
CY	3,355
EE	4,738
ES	12,866
GR	6,499
IE	5,243
IT	20,925
LU	3,755
LV	5,192
UK	8,850

Source: EU-SILC 2008 cross-sectional data-files, version UDB 01.08.10 Authors' calculation.
The unit of analysis is the household

Table A2. Citizenship of the respondents by country (2008)

Country	Local	Other
AT	92.22	7.78
BE	92.91	7.09
CY	89.84	10.16
EE	84.1	15.9
ES	94.67	5.33
GR	94.15	5.85
IE	92.75	7.25
IT	94.76	5.24
LU	61.5	38.5
LV	81.44	18.56
UK	95.14	4.86

Source: EU-SILC 2008 cross-sectional data-files, version UDB 01.08.10 Authors' calculation.
The unit of analysis is the household.

CEPS
INSTEAD

3, avenue de la Fonte
L-4364 Esch-sur-Alzette
Tél.: +352 58.58.55-801
www.ceps.lu