Socioeconomic Status and Health: 
A New Approach to the Measurement of Bivariate Inequality

Guido Erreygers & Roselinde Kessels
FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ECONOMICS

Socioeconomic Status and Health: A New Approach to the Measurement of Bivariate Inequality

Guido Erreygers & Roselinde Kessels

RESEARCH PAPER 2015-017
AUGUST 2015

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@uantwerpen.be

The research papers from the Faculty of Applied Economics are also available at www.repec.org
(Research Papers in Economics - RePEc)

D/2015/1169/017
Socioeconomic Status and Health: A New Approach to the Measurement of Bivariate Inequality

Guido Erreygers and Roselinde Kessels

August 17, 2015

Abstract

We suggest an alternative way to construct a family of indices of socioeconomic inequality of health. Our indices belong to the broad category of linear indices. In contrast to rank-dependent indices, which are defined in terms of the ranks of the socioeconomic variable and the levels of the health variable, our indices are based on the levels of both the socioeconomic and the health variable. We also indicate how the indices can be modified in order to introduce sensitivity to inequality in the socioeconomic distribution and to inequality in the health distribution. As an empirical illustration, we make a comparative study of the relation between income and well-being in 16 European countries using data from SHARE Wave 4.

Keywords: Inequality measurement, Socioeconomic inequality of health, Bivariate inequality.

JEL Classification Number: D63, I00

Affiliation of the authors:
Guido Erreygers: Department of Economics, University of Antwerp, Belgium & Centre for Health Policy, University of Melbourne, Australia
guido.erreygers@uantwerpen.be
Roselinde Kessels: Department of Economics, University of Antwerp & Flemish Research Foundation (FWO), Belgium
roselinde.kessels@uantwerpen.be

Acknowledgements: We are grateful to Karel Van den Bosch for his help with regard to the SHARE data.
1 Introduction

Inequalities in socioeconomic outcomes or capabilities (income, wealth, etc.), especially when they are large, are sometimes perceived as excessive and unfair. It might be argued, for instance, that to the extent that these inequalities are undeserved, they are unjustified and should be eliminated. Another reason why people object to severe socioeconomic inequalities is that they tend to spread to other domains. This concern certainly provides one of the motivations for Michael Walzer’s notion of complex equality:

The regime of complex equality is the opposite of tyranny. It establishes a set of relationships such that domination is impossible. In formal terms, complex equality means that no citizen’s standing in one sphere or with regard to one social good can be undercut by his standing in some other sphere, with regard to some other good. (Walzer, 1983: 19)

Put differently, socioeconomic inequality appears to be more worrisome if rich people, besides having higher incomes than poor people, also have systematically better access to education, political power, health care, and so on. There is a lot of evidence which suggests that this is indeed the case. The huge literature on the social gradient in health amply illustrates that inequality in socioeconomic conditions is often positively and strongly associated to inequality in health achievements (see, e.g., Adler et al., 1994, Marmot et al., 2010, and many others).

This paper focuses on the measurement of the extent to which socioeconomic status and health are related to one another. Rank-dependent indicators are by far the most widely used indices to measure socioeconomic inequality of health. The existing rank-dependent indices have been developed to help us find out whether there is pro-rich or pro-poor bias in the health distribution: positive values indicate that people who are relatively well-off in socioeconomic terms tend to have better health than those who are less well-off, and negative values the opposite. The Concentration Index (Wagstaff, Paci and Van Doorslaer, 1991), the extended Concentration Index (Pereira, 1998, and Wagstaff, 2002), the ‘bounded variable’ indices proposed by Wagstaff (2005) and Erreygers (2009), and even the Symmetric Index (Erreygers, Clarke and Van Ourti, 2012) are all indices of this type.

What characterizes rank-dependent indicators is that they measure socioeconomic status by the ranks which individuals occupy in the socioeconomic distribution, but health by the levels of the health variable under consideration. It could be argued that this asymmetric treatment of socioeconomic status and health reflects a causal relationship between socioeconomic
ranks and health levels. The ‘relative position’ hypothesis, however, is only one of several possible explanations of the association between socioeconomic status and health (see, e.g., O’Donnell, Van Doorslaer and Van Ours, 2015, section 5.3). In the absence of any clear evidence on the exact nature of the link between socioeconomic status and health, the choice seems unduly restrictive. There are also no compelling empirical reasons for the asymmetry, because in most cases we do have information on the exact socioeconomic levels which people achieve, e.g., their incomes. Yet rank-dependent indices do not use that information; they simply rely on the ranks, not the levels.

Our main aim in this paper is to propose a set of indices based on the levels which individuals attain in the health and socioeconomic domains. We believe that indices based on socioeconomic levels rather than ranks give a more complete picture of the socioeconomic distribution, and therefore provide a more accurate measure of socioeconomic inequality of health. A related objective of our paper is to make room for distributional sensitivity. One may have good reasons to be relatively more concerned about those who are not very well-off in socioeconomic terms, or who are in particularly bad health. In the literature so far, the focus has been on sensitivity to inequality in the socioeconomic distribution. The extended Concentration Index, for instance, has been constructed by applying suggestions made by Kakwani (1980), Donaldson and Weymark (1980) and Yitzhaki (1983) in the context of income inequality measurement, to rank-dependent indices of socioeconomic inequality. The indices we suggest here can be made sensitive to both the socioeconomic and the health distribution.

2 Indices Based on Socioeconomic Ranks

2.1 Preliminaries

Suppose we have information on the socioeconomic achievement \((y_i)\) and the health attainment \((h_i)\) of each individual \(i\) in a population consisting of \(n\) individuals \((i = 1, 2, ..., n)\). Assume that both \(y_i\) and \(h_i\) have well-defined lower bounds greater than or equal to 0. More specifically, each of these variables is either a ratio-scale variable with no upper bound, or a cardinal or ratio-scale variable with a well-defined upper bound. The socioeconomic and health variables are not necessarily of the same nature. Typically, the socioeconomic variable \(y\) is an unbounded ratio-scale variable, e.g. income or consumption; this is also the assumption adopted in our paper. The health variable, by contrast, is often a bounded variable, and need not be of the ratio-scale type. The means of the variables are \(\mu_y = \frac{1}{n} \sum_{i=1}^{n} y_i\) and
\( \mu_h = \frac{1}{n} \sum_{i=1}^{n} h_i \), respectively.

Given the levels of the socioeconomic and health variables, we can determine the ranks of all individuals in the socioeconomic and health distributions. We denote the rank of individual \( i \) in the socioeconomic distribution as \( r_i \) and in the health distribution as \( q_i \). If there are no ties in a distribution, the rank of the person with the lowest level is equal to 1, the rank of the person with the second lowest level equal to 2, etc., and the rank of the person with the highest level equal to \( n \). If a group of \( k+1 \) individuals are tied on position \( g \), the rank of each of these individuals is equal to \( g + (k/2) \). For simplicity, however, we assume there are no ties in the distributions. This assumption is made in order to simplify the expressions of the rank-dependent indices; it is by no means necessary.

### 2.2 Bivariate Linear Indices

The rank-dependent indices used in the literature on the measurement of socioeconomic inequality of health can be seen as belonging to a broad class of inequality measures. These indices are closely related to the family of linear measures of inequality defined by Mehran (1976). Mehran focused on the measurement of univariate inequality, say inequality in distribution \( x \). Following the representation of positional univariate indices suggested by Lambert and Lanza (2006: 260-261), we can express the absolute version of the family of univariate linear measures of inequality for discrete distributions as a weighted average of the \( x_i \) levels:

\[
U_A(x) = \frac{1}{n} \sum_{i=1}^{n} w_i(x) x_i
\]

where \( w_i(x) \) is the weight of individual \( i \), determined in some way by distribution \( x \). Typical for rank-dependent measures is that they are characterized by weights which are defined in terms of the ranks of individuals in distribution \( x \). But, obviously, one can think of other ways of linking the weights to distribution \( x \). In any case, in order to guarantee that a perfectly equal distribution of \( x \) leads to a zero degree of inequality, these weights must be such that:

\[
\sum_{i=1}^{n} w_i(x) = 0
\]

whenever \( x \) is perfectly equally distributed. Additional properties such as transfer sensitivity impose further conditions on the weights.

In this paper we focus on bivariate rather than univariate inequality, and therefore we have to give a slightly different interpretation to the formulas.
In the bivariate case, the weights are determined by a different distribution from the one which is weighted. In other terms, the weights are linked to the distribution of a weighting variable which is different from the weighted variable. The Concentration Index, for example, defines the weights in terms of the ranks of individuals in the socioeconomic distribution, and applies these weights to the distribution of health. In formal terms, the absolute version of the bivariate linear index where \( z \) serves as the weighting variable and \( x \) as the weighted variable can be defined as:

\[
B_A(z, x) = \frac{1}{n} \sum_{i=1}^{n} w_i(z)x_i
\]  

(3)

The relative versions of the univariate and bivariate indices are obtained by dividing the index by the mean of the weighted variable. The relative version of the univariate index is therefore \( U_R(x) = \frac{1}{\mu_x} U_A(x) \), and that of the bivariate index \( B_R(z, x) = \frac{1}{\mu_x} B_A(z, x) \). Other versions of the indices have been proposed in the literature, e.g. in order to deal with bounded variables (Wagstaff, 2005; Erreygers, 2009). Erreygers and Van Ourti (2011) have argued that the choice for a specific version of a bivariate index should be made in accordance with the type of weighted variable under consideration.

We will now explore what kind of conditions can reasonably be imposed upon the weights \( w_i(z) \).

2.3 Properties of Bivariate Linear Indices

Bivariate indices of inequality serve another purpose than univariate indices of inequality. Bivariate indices measure the degree of correlation between two distributions: in statistical terms, they focus on the covariance between two variables rather than on the variance of one variable. When examining whether a particular bivariate linear index is acceptable or not, we therefore have to use a checklist of desirable properties which relate to the correlation between the variables under consideration. In this section we enounce three properties and specify the tests we will use to check whether the properties hold.

2.3.1 Absence of correlation

The first property refers to how an index evaluates situations in which the two variables are uncorrelated.

**Zero correlation:** If there is no correlation between the weighting and the weighted variable, the measured degree of bivariate inequality must be zero.
The main function of the Zero correlation property is to define the borderline between cases of positive correlation, indicated by positive values of the index, and cases of negative correlation, indicated by negative values. With regard to socioeconomic inequality of health, a positive correlation means that there is a pro-rich bias in the distribution of health, and a negative correlation that there is a pro-poor bias.

There are two elementary tests we can use to check whether this property holds. We can in fact identify two special cases of no correlation: the first occurs when the weighted variable is a constant no matter what the level of the weighting variable is, and the second when the weighting variable is a constant no matter what the level of the weighted variable is. This means that the index must be zero both when there is perfect equality of the weighted variable and when there is perfect equality of the weighting variable. The first test leads to the condition that the sum of the weights must be zero:

$$\sum_{i=1}^{n} w_i(z) = 0$$

(4)

The second test imposes a weaker condition: it suffices that condition (4) holds whenever there is perfect equality in distribution $z$. Since it seems obvious that perfect equality in distribution $z$ implies equal weights for all individuals, all the weights $w_i(z)$ must equal 0.

We consider only bivariate linear indices which satisfy condition (4), and which therefore pass the two elementary tests. But we can go one step further. A slightly more demanding version of the two tests consists of checking whether the measured degree of bivariate inequality tends to zero when the distribution of either $x$ or $z$ approaches that of perfect equality. This implies in particular that when the distribution of the weighting variable $z$ tends to perfect equality, the weights $w_i(z)$ must change and tend to 0.

### 2.3.2 Changes in correlation

Next, we specify how the index should react to changes in correlation.

**Correlation sensitivity:** Any change which increases the positive (c.q. negative) correlation between the weighting and the weighted variable, must lead to a positive (c.q. negative) change of the measured degree of bivariate inequality.

We use the following simple tests to check whether this property holds. Suppose there is a ‘transfer’ from one person to another in the distribution of the weighted variable which is to the advantage of the person who has the highest level of the weighting variable. (If health is the weighted variable and
socioeconomic status the weighting variable, this means a ‘transfer’ of health from a relatively worse-off person to a relatively better-off person.) Since this increases the positive correlation between the two variables, the value of the index should increase. Likewise, if there is a ‘transfer’ from one person to another in the distribution of the weighting variable which happens to be to the advantage of the person who has the highest level of the weighted variable, the value of the index should increase.

It is relatively easy to determine the condition which the first test imposes upon the weights. The test involves a change in distribution $x$ only. Suppose there is a transfer $\delta_x > 0$ from individual $i$ to individual $j$, with individual $i$ worse off than individual $j$ in terms of the weighting variable ($z_i < z_j$). As a result of this, the level of individual $i$ changes from $x_i$ to $x_i - \delta_x$ and that of individual $j$ from $x_j$ to $x_j + \delta_x$, while everything else remains the same. The change in the value of the index $B_A(z, x)$ is therefore equal to $\delta_x [w_j(z) - w_i(z)]$. Obviously, the value of the index increases if and only if $w_j(z) - w_i(z) > 0$. This means that the weights must be increasing in the weighting variable, i.e. $w_i(z) < w_j(z)$ if and only if $z_i < z_j$.

The implications of the second test are a bit more complicated. The problem is that a transfer occurring in distribution $z$ can modify the weights of more individuals than the ones who are affected by the transfer. To begin with, suppose there is a transfer $\delta_z > 0$ from individual $i$ to individual $j$, with individual $i$ worse off than individual $j$ in terms of the weighted variable ($x_i < x_j$), which changes only the weights of individuals $i$ and $j$. Let the differences in the weights be $\Delta w_i$ and $\Delta w_j$, which of course must be such that $\Delta w_i + \Delta w_j = 0$. The change in the value of the index $B_A(z, x)$ is then equal to $\Delta w_i x_i + \Delta w_j x_j = \Delta w_j (x_j - x_i)$. Obviously, the change is positive if and only if $\Delta w_j > 0$. This means that the weight of the individual whose position improves as a result of the transfer $\delta_z$ must increase, which boils down to the condition that the weights must be increasing in the weighting variable. However, if the transfer affects the weights of other individuals besides $i$ and $j$, we have a different situation. Suppose the transfer affects also the weight of individual $k$, while all other weights remain the same. We now have $\Delta w_i + \Delta w_j + \Delta w_k = 0$, and the change in the value of the index $B_A(z, x)$ is equal to $\Delta w_i x_i + \Delta w_j x_j + \Delta w_k x_k = \Delta w_i (x_i - x_k) + \Delta w_j (x_j - x_k)$. Since the signs of $(x_i - x_k)$ and $(x_j - x_k)$ are undetermined, we are unable to impose conditions on $\Delta w_i$ and $\Delta w_j$ which ensure that the change is positive. Hence, we arrive at the conclusion that if a transfer in the distribution of the weighting variable affects only the weights of the individuals concerned by the transfer, and if moreover the weights are increasing in the weighting variable, then Correlation sensitivity holds. However, if the transfer affects the weights of other individuals as well, it is uncertain whether Correlation
sensitivity holds.

### 2.3.3 Distributional sensitivity

For those indices which are sensitive to changes in correlation, one might ask to what extent they are sensitive to transfers occurring at different locations in the distribution. Two cases must be distinguished. Suppose we are looking at the effect of a transfer in the distribution of the weighted variable $x$ between different pairs of people who are all at equal distance in the distribution $z$. Given that a transfer $\delta_x$ between persons $i$ and $j$ changes the value of the index by $\delta_x [w_j(z) - w_i(z)]$, the magnitude of the effect is proportional to the difference between the weights assigned to $i$ and $j$. If this difference depends on the location of $i$ and $j$ in the distribution of the weighting variable $z$, the effect of the transfer $\delta_x$ will be sensitive to the distribution of $z$. Suppose, alternatively, that we are looking at the effect of a transfer in the distribution of the weighting variable $z$ between different pairs of people who are all at equal distance in the distribution $x$. Let us consider the case where a transfer $\delta_z$ between persons $i$ and $j$ changes the index by the amount $\Delta w_j(x_j - x_i)$. The magnitude of the effect is proportional to the difference between the levels attained by $i$ and $j$ in distribution $x$. This means that the effect of $\delta_z$ is the same for people at equal distance in the distribution of $x$, and therefore that the effect of the transfer $\delta_z$ will be insensitive to the distribution of $x$. It follows that we can only talk about sensitivity to transfers in the distribution of the weighted variable occurring at different locations in the distribution of the weighting variable.

In line with what is customary in the literature on income inequality, the prevalent notion is to give more weight to changes occurring at the lower end of the distribution. This leads to the formulation of the following property.

**Lower-end distributional sensitivity:** A given change in the distribution of the weighted variable has a larger absolute impact when it occurs at the lower end of the distribution of the weighting variable than when it occurs at the higher end.

The implication of this property can be tested by looking at the effect of a given transfer $\delta_x$ taking place among equidistant individuals at different locations in the distribution $z$. Let us compare the effect of a transfer $\delta_x > 0$ between persons $i$ and $j$ with that of the same transfer $\delta_x$ between persons $k$ and $l$, where $z_i - z_j = z_k - z_l < 0$ and $z_i < z_k$. In the first case, the change of the index is $\delta_x [w_j(z) - w_i(z)]$ and in the second $\delta_x [w_l(z) - w_k(z)]$. There is lower-end distributional sensitivity if and only if $w_j(z) - w_i(z) > w_l(z) - w_k(z)$. In other terms, the absolute difference between the weights of equidistant individuals must become smaller and smaller as one moves from
the low end to the high end of distribution \( z \). Assuming that the weights are increasing in the weighting variable, this means that the rate at which the weights increase becomes smaller and smaller.

Distributional sensitivity need not be limited to the lower end of the distribution. Erreygers, Clarke and Van Oursi (2012) have made a case for sensitivity to changes occurring at both ends of the distribution. However, what they have called the symmetry property can be defined only if the weighting variable follows a uniform distribution. Since this is in general not the case, we will limit ourselves to the case of lower-end distributional sensitivity.

2.4 Rank-Dependent Indices

Rank-dependent indices are special cases of linear indices. Given a weighting variable \( z \), such as income or socioeconomic status, they use the rank of person \( i \) in distribution \( z \) to determine the weight of person \( i \). The level of person \( i \) in distribution \( z \) serves only to derive the rank of that person; it does not enter directly into the calculation of the weight. The standard Concentration Index, for instance, is characterized by a weighting function which is linear in the ranks \( r_i \) determined by the socioeconomic distribution \( y \). The weighting function may therefore be called the ‘socioeconomic rank’ function and the associated index the \( SR \) index:

\[
w_i(y) = w_i^{SR} = \frac{2r_i - n - 1}{n}
\]  

The weights steadily increase as \( r_i \) goes from 1 to \( n \). If we define the group of the ‘poor’ as those who have negative weights and the group of the ‘rich’ as those who have positive weights, the weighting function (5) puts the boundary between the two groups exactly in the middle of the population. Those with ranks smaller than or equal to \( n/2 \) (if \( n \) is even) or smaller than or equal to \( (n-1)/2 \) (if \( n \) is odd) have negative weights, and those with ranks larger than or equal to \( n/2 + 1 \) (if \( n \) is even) or larger than or equal to \( (n+1)/2 \) (if \( n \) is odd) have positive weights.

There is no compelling reason why rank-dependent indices must be based on a linear weighting function such as (5). The extended Concentration Index has been developed to give relatively more weight to individuals with lower ranks. Slightly generalizing the approach of Erreygers, Clarke and Van Oursi (2012), the extended version of the \( SR \) weighting function can be expressed
as follows:

$$w^R_i(\nu) = \begin{cases} 
1 + n \left(\frac{n-r_i}{n}\right)^\nu - \left(\frac{n-r_i+1}{n}\right)^\nu, & \nu > 0, \nu \neq 1 \\
\nu - 1 & 
\end{cases}$$

$$\left(\frac{n-r_i}{n}\right) \ln \left(\frac{n-r_i}{n}\right) - \left(n - r_i + 1\right) \ln \left(\frac{n-r_i+1}{n}\right), \quad \nu = 1$$

where $\nu$ is a distributional sensitivity parameter. The extended $SR$ index is denoted by $SR(\nu)$. The linear weighting function (5) corresponds to the value $\nu = 2$. The higher the value of $\nu$, the more sensitive the index is to changes in the lower end of the rank distribution. The sum of the negative weights is equal to $-n/(\nu^{\nu/(\nu-1)})$, and that of the positive weights $n/(\nu^{\nu/(\nu-1)})$. For $\nu > 1$, the lowest negative weight is (approximately) $-1$, and the highest positive weight (approximately) $1/(\nu - 1)$.

### 2.5 Assessing Rank-Dependent Indices

The main drawback of rank-dependent indices is that a lot of information goes lost by focusing only on the ranks of the weighting variable. Strictly speaking, rank-dependent indices capture the correlation between the ranks of the weighting variable and the levels of the weighted variable rather than the correlation between the levels of the weighting variable and the levels of the weighted variable. Although rank-dependent indices pass the two elementary tests of the Zero correlation property, they fail the more demanding version of the two tests. Starting from an unequal distribution $y$, imagine that we change everyone’s position from $y_i$ to $\gamma y_i + (1-\gamma)\mu y_i$, where $0 \leq \gamma \leq 1$. The smaller the value of $\gamma$, the more we tend to a situation of perfect equality. Yet, as we move from $\gamma = 1$ (i.e., the original situation) to $\gamma = 0$ (i.e., the situation of perfect equality), all ranks remain the same, except at the moment when $\gamma = 0$. Hence, rank-dependent indices do not tend to zero when the distribution of the weighting variable approaches perfect equality; instead, they remain constant as long as $\gamma > 0$, and suddenly jump to zero when $\gamma = 0$. More generally, changes in the levels of the weighting variable which do not lead to changes in the ranks of individuals have no effect at all on the value of rank-dependent indices. In formal terms, this means that rank-dependent indices do not have the property of Correlation sensitivity. Since they do not always react to changes in the distribution of the weighting variable, it cannot be predicted how sensitive they are to transfers occurring at the lower or higher ends of this distribution. Hence, they also do not have the property of Lower-end distributional sensitivity.

What lies at the root of the problem is the fact that the ranks of individuals say very little about the levels which these individuals attain, and
that the difference between the ranks of two individuals conveys almost no information about the difference in the levels which these individuals attain. The levels and their differences may be very big, very small, or something in between. The challenge, therefore, is to find a way to construct an indicator which does take the levels of the weighting variable into account in the determination of the weights. Put differently, we want to explore whether and how an indicator can be based on level-dependent rather than rank-dependent weights.

3 Indices Based on Socioeconomic Levels

3.1 The Basic Construction

We now introduce an alternative way to define the individual weights. Instead of letting the weights depend exclusively on the ranks of individuals, we define them in terms of the levels of the socioeconomic variable. In other words, we work with a weighting function, the arguments of which are levels, not ranks. What we are looking for is a set of individual weights which satisfy the condition that their sum must equal zero, as required by the property of Zero correlation. Moreover, the weights should be such that the indices have the property of Correlation sensitivity, and if possible that of Lower-end distributional sensitivity. We have already established that Correlation sensitivity means that the weights must be increasing in the socioeconomic levels, and Lower-end distributional sensitivity that the differences between the weights must become smaller as one moves up the socioeconomic ladder.

There are undoubtedly many ways in which such a set of weights can be defined. A simple procedure consists of making an individual’s weight proportional to the deviation of this person’s socioeconomic level $y_i$ from the mean socioeconomic level $\mu_y$, i.e. proportional to $d_i = y_i - \mu_y$. Since $d_i < d_j \iff y_i < y_j$ and $\sum_{i=1}^{n} d_i = 0$, this ensures that the weights are increasing in the socioeconomic levels and sum to zero. In order to make the weights unit-free (e.g., if $y$ is income, we would like the weights to be independent of the chosen monetary unit), some kind of normalization must be adopted. One way of obtaining unit-free weights is to define the factor of proportionality in terms of the absolute mean deviation $\mu_{|d|}$, where $\mu_{|d|} = \frac{1}{n} \sum_{i=1}^{n} |d_i|$. More specifically, suppose we define the weight $w_i(y)$ as the ratio of the deviation of $y_i$ from $\mu_y$ to the absolute mean deviation:

$$w_i(y) = \frac{d_i}{\mu_{|d|}} = \frac{y_i - \mu_y}{\frac{1}{n} \sum_{j=1}^{n} |y_j - \mu_y|}$$
Given that $\sum_{i=1}^{n} d_i = 0$, the sum of the positive deviations equals the absolute value of the sum of the negative deviations. Since both of these amount to $n\mu_{|\bar{d}|}/2$, this specific choice of the proportionality factor would ensure that all the negative weights sum to $-n/2$, and all the positive weights to $n/2$.

We cannot stop there, however. Let us imagine that all $y_i$ are changed into $\tilde{y}_i = \gamma y_i + (1 - \gamma) \mu_y$, where $0 \leq \gamma \leq 1$. Since $\mu_{\tilde{y}} = \mu_y$, it follows that $\tilde{d}_i = \tilde{y}_i - \mu_{\tilde{y}} = \gamma d_i$. This means that the $\tilde{y}$ distribution is definitely more equal than the $y$ distribution: all distances to $\mu_y$ have become smaller. As a result, one would expect that the weights decrease in absolute value. But since $\mu_{|\tilde{d}|} = \gamma \mu_{|d|}$, this would not happen if we used (7) to define the weights. Given that $\tilde{d}_i/\mu_{|\tilde{d}|} = d_i/\mu_{|d|}$, the new weights would be exactly the same as the old ones.

We therefore suggest to correct (7) by multiplying these weights by a scalar measuring the degree of inequality of distribution $y$. An obvious choice would be the Gini coefficient. This measure, however, is of the rank-dependent type, and we wish to move away from rank-dependency. We therefore opt for the relative mean deviation, which is equal to the absolute mean deviation, $\frac{1}{n} \sum_{i=1}^{n} |y_i - \mu_y|$, divided by the mean, $\mu_y$.\footnote{This is acceptable only if $y$ has ratio-scale properties. If $y$ is a cardinal rather than a ratio-scale variable, we have to choose an inequality measure which is invariant to any transformation $a + by$, with $b > 0$.} Then the weighting function of the socioeconomic level-dependent index $SL$ becomes:

$$w_{iSL} = \frac{y_i - \mu_y}{\frac{1}{n} \sum_{j=1}^{n} |y_j - \mu_y|} = \frac{y_i - \mu_y}{\mu_y}$$

(8)

In the scenario we have just considered, the new distribution $\tilde{y}$ is certainly more equal than the old distribution $y$. Every weight decreases in absolute value and is equal to $\gamma$ times its original value. Note that all the negative weights now sum to $-(n/2)(\mu_{|d|}/\mu_y)$, and all the positive weights to $(n/2)(\mu_{|d|}/\mu_y)$.

### 3.2 A Parametric Extension

The weights (8) are dependent both upon the levels (not the ranks) of the socioeconomic variable and upon the degree of inequality in the socioeconomic distribution. The deviations of the socioeconomic levels from the mean determine the relative positions of the individuals, while the degree of inequality fixes the bounds of the weights. If there is a high degree of inequality in the socioeconomic distribution, typically there will be a lot of individuals with
socioeconomic levels below the mean, who will therefore have negative, but in absolute terms rather small weights. On the other side of the spectrum, those who are well-off in socioeconomic terms will have positive, and in absolute terms quite large weights. Hence, a small change in the health level of a relatively well-off individual will probably have a more pronounced influence on the index of socioeconomic inequality than a comparable change in the health level of an individual who is less well-off. This is perhaps not the kind of sensitivity which we want our index to have.

One way of making the weights more sensitive to the bottom of the socioeconomic distribution consists of first applying a transformation to the socioeconomic levels. A convenient parametric form is of the isoelastic type:

$$y_i(\alpha) = \begin{cases} \frac{y_i^{1-\alpha} - \alpha}{1-\alpha} & (\alpha \neq 1) \\ 1 + \log(y_i) & (\alpha = 1) \end{cases}$$  

(9)

(In order to make sure that $y_i(\alpha)$ exists for $\alpha \geq 1$, we assume that all socioeconomic levels $y_i$ are strictly positive.)

We can likewise define $\mu_y(\alpha) = \frac{1}{n} \sum_{i=1}^{n} y_i(\alpha)$. Let us then define the relative position of individual $i$ as the ratio of the deviation $d_i(\alpha) = y_i(\alpha) - \mu_y(\alpha)$ to the absolute mean deviation $\mu_y|d|$, defined as $\mu_y|d| = \frac{1}{n} \sum_{i=1}^{n} |y_i(\alpha) - \mu_y(\alpha)|$. Keeping the inequality component unchanged, we arrive at the following definition of the weights:

$$w_{i}^{SL}(\alpha) = \frac{y_i(\alpha) - \mu_y(\alpha)}{\frac{1}{n} \sum_{j=1}^{n} |y_j(\alpha) - \mu_y|} \frac{1}{\mu_y} \sum_{j=1}^{n} |y_j - \mu_y|$$  

(10)

with the corresponding bivariate inequality index denoted by $SL(\alpha)$.

It is easy to see that the case $\alpha = 0$ corresponds to our original proposition: no transformation is applied to the socioeconomic levels. As $\alpha$ increases, the weight of the most well-off individual remains positive but decreases in magnitude, whereas the weight of the least well-off individual remains negative but increases in magnitude. What happens to the weights of the other individuals depends on the specific socioeconomic distribution. More and more individuals who initially had a negative weight will get a positive weight, until eventually only the least well-off individual will be the only one with a negative weight. As in the basic case, all the negative weights sum to $-(n/2)(\mu_{d}/\mu_{y})$, and all the positive weights to $(n/2)(\mu_{d}/\mu_{y})$. Since $\mu_{d}$ is at most equal to $2(n-1)\mu_{y}/n$, it follows that these two amounts are

---

2This excludes the use of data according to which some individuals attain a zero or negative level of socioeconomic achievement. This happens frequently when income data are used for the measurement of socioeconomic status. Instead of income, one might consider using consumption data.
bounded by $1 - n$ and $n - 1$. For high values of $\alpha$, therefore, the lowest negative weight will tend to $-(n/2)(\mu_{|d|}/\mu_y) \geq 1 - n$, while all the other weights will be positive and tend to $(n/2)(\mu_{|d|}/\mu_y)/(n - 1) \leq 1$.

For a given value of $\alpha$, a change of unit of the socioeconomic variable (e.g., from € to $) has no effect on the weights. As a matter of fact, all $y_i(\alpha)$, and therefore also $\mu_y(\alpha)$, change in the same proportion.

3.3 Assessing Level-Dependent Indices

It remains to be seen how the level-dependent indices $SL$ and $SL(\alpha)$ perform on the tests we have proposed for the three properties of Zero correlation, Correlation sensitivity and Lower-end distributional sensitivity. Obviously, these indices pass the Zero correlation tests, since their weights sum to zero, and moreover the weights decrease in absolute value and tend to zero when the distribution of $y$ approaches perfect equality. The $SL$ index passes the Correlation sensitivity test: a transfer $\delta_y$ between two individuals affects only the weights of these individuals, and in addition the weights are increasing in the socioeconomic variable. However, since the $SL$ weighting function is linear in this variable, the index does not pass the Lower-end distributional sensitivity test. This was the reason why we introduced the $SL(\alpha)$ index, with $\alpha > 0$. The nonlinear character of the $SL(\alpha)$ weighting function ensures that this index passes the Lower-end distributional sensitivity test. But that comes at a price: the $SL(\alpha)$ index no longer passes the Correlation sensitivity test. To be precise, it may happen that a transfer $\delta_y$ which increases the correlation between the socioeconomic and the health variable does not lead to an increase of the index. The reason is that for $\alpha > 0$, a transfer $\delta_y > 0$ from individual $i$ to individual $j$ changes the weights of all individuals, not just those of $i$ and $j$. The weight of individual $i$ decreases and that of individual $j$ increases, while the weights of other individuals may either increase or decrease. This is due to the fact that $\mu_y(\alpha)$ decreases, which changes all deviations $d_i(\alpha) = y_i(\alpha) - \mu_y(\alpha)$. As a result, the absolute mean deviation $\mu_{|d(\alpha)|}$ also changes. Since it is possible that $\mu_{|d|}$ also changes, we have three factors which may affect the values of the weights.$^3$

Even if we ignore the effects of the changes in the other weights, which will be small, we cannot be sure that the $SL(\alpha)$ index changes in the right direction. In order to pass the Correlation sensitivity test, the value of $SL(\alpha)$ should increase in reaction to a transfer $\delta_y > 0$ from individual $i$ to individual $j$, where $j$ is in better health than $i$ ($h_i < h_j$). It could very well be that the decrease of $w_i$ is larger than the increase of $w_j$, i.e., that

$^3$The weighting function can in fact be written as: $w_{SL}^i(\alpha) = (d_i(\alpha)/\mu_{|d(\alpha)|}) (\mu_{|d|}/\mu_y)$. 

14
\(\Delta w_i + \Delta w_j < 0\). In that case, the index will decrease rather than increase if we have \(-\Delta w_j/\Delta w_i < h_i/h_j\). It seems that only for relatively high values of \(\alpha\), \(\Delta w_j\) will be substantially smaller than \(-\Delta w_i\). We can expect, therefore, that for relatively small values of \(\alpha\), the index \(SL(\alpha)\) will nearly always change in the right direction in response to a transfer \(\delta_y\). In any case, the index always reacts appropriately to a transfer \(\delta_h\) of the health variable. Hence, the price we have to pay for the introduction of Lower-end distributional sensitivity seems relatively modest.

4 Indices Based on Health Ranks or Levels

4.1 An Inverted Approach

In the previous section, we followed the usual practice of treating socio-economic status as the weighting variable, and health as the weighted variable. We have indicated how the weights can be defined in terms of the socioeconomic levels rather than the ranks, and how these weights can be made dependent upon the socioeconomic distribution in order to express concern for the poorer part of the population. But what if we want to introduce distributional sensitivity to the lower end of the health distribution and express concern for those who are worse-off in terms of health?

In our opinion, the easiest solution consists of turning the approach on its head. Let us return to the case of rank-dependent indices. It is well-known that the Concentration Index has been developed by analogy to the Gini coefficient. The Gini coefficient is a univariate inequality measure which uses both the levels and the ranks of the single variable of interest (e.g., income). The Concentration Index, by contrast, is a bivariate inequality measure which considers the ranks of the socioeconomic variable and the levels of the health variable. However, it is nowhere carved in stone that the socioeconomic variable should enter only through its ranks and the health variable only through its levels, and not the other way around. As a matter of fact, if the health variable is of an ordinal kind, it seems more natural to use the ranks of the health variable and the levels of the socioeconomic variable (see Erreygers, 2009). This leads to a family of indices in which the weighted variable is socioeconomic status and the weights depend on the ranks which individuals occupy in the health distribution. The counterpart of the health Concentration Index would then be the socioeconomic status Concentration Index. It is based on the ‘health rank’ weighting function:

\[
w_i^{HR} = \frac{2q_i - n - 1}{n}
\]  

(11)
where \( q_i \) represents the rank of individual \( i \) in the health distribution. The associated bivariate inequality index is denoted by \( HR \). Likewise, the extended version of this index, \( HR(\lambda) \), would be based on the following weighting function:

\[
w_i^{HR}(\lambda) = \begin{cases} 
1 + n \left[ \left( \frac{n-q_i}{n} \right)^\lambda - \left( \frac{n-q_i+1}{n} \right)^\lambda \right], & \lambda > 0, \lambda \neq 1 \\
(n - q_i) \ln \left( \frac{n-q_i}{n} \right) - (n - q_i + 1) \ln \left( \frac{n-q_i+1}{n} \right), & \lambda = 1
\end{cases}
\] (12)

with \( \lambda \) as the distributional sensitivity parameter.

4.2 Health Levels Instead of Health Ranks

As before, we can raise the objection that if we focus on the health ranks only, we throw away a lot of relevant information. Hence, it seems useful to define the weights in terms of the health levels.

Following the same reasoning as before, a first possibility is that we use the following level-dependent weighting function:

\[
w_i^{HL} = \frac{h_i - \mu_h}{\mu_h}
\] (13)

which leads to the index \( HL \). If we want to introduce sensitivity to the health distribution, we can do so by first transforming the health levels:

\[
h_i(\beta) = \begin{cases} 
h_i^{1-\beta} & (\beta \neq 1) \\
1 + \log(h_i) & (\beta = 1)
\end{cases}
\] (14)

and then using the following weighting function:

\[
w_i^{HL}(\beta) = \frac{\frac{1}{n} \sum_{j=1}^{n} |h_j - \mu_h|}{\frac{1}{n} \sum_{j=1}^{n} |h_j(\beta) - \mu_h(\beta)|}
\] (15)

This yields the index \( HL(\beta) \). Higher values of \( \beta \) make the index more sensitive to the lower end of the health distribution.

5 Rank-Dependent vs. Level-Dependent Indices

5.1 A Comparison of the Indices

It may be useful to sum up how we suggest to move from rank-dependent indices to level-dependent indices. For each rank-dependent index we have constructed a level-dependent alternative, as indicated here below:
If the levels of both the socioeconomic status and the health variable are known, there is in general no reason why one should stick to rank-dependent indices. Only when some of the levels of the weighting variable are equal to zero, the extended versions of the level-dependent indices cannot be computed for \( \alpha, \beta \geq 1 \).

The basic versions \( SR, SL, HR \) and \( HL \) are all special cases of the extended versions, taking \( \nu, \lambda = 2 \) and \( \alpha, \beta = 0 \). Since rank-dependent weights differ from level-dependent weights (except by fluke), the values of the associated indices are also different. Put differently, in general \( SR \) is unrelated to \( SL \), and \( HR \) to \( HL \). Likewise, there is no reason why the two rank-dependent indices \( SR \) and \( HR \) should be related to one another. But the same does not hold with respect to the two level-dependent indices. This can be illustrated by expressing \( SL \) and \( HL \) in terms of covariances. In fact, we have:

\[
SL = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \mu_y \right) h_i = \frac{Cov(y, h)}{\mu_y} \tag{16}
\]

\[
HL = \frac{1}{n} \sum_{i=1}^{n} \left( h_i - \mu_h \right) y_i = \frac{Cov(y, h)}{\mu_h} \tag{17}
\]

From these expressions we can derive that the relative versions of the two indices are equal to one another:

\[
\frac{SL}{\mu_h} = \frac{HL}{\mu_y} = \frac{Cov(y, h)}{\mu_y \mu_h} \tag{18}
\]

Expression (18) reveals that the two basic level-dependent indices are closely related to the Atkinson-type index derived by Erreygers (2013), and to the class of covariance-based indices derived axiomatically by Bidard (2009).

### 5.2 A Comparison of the Weighting Functions

We can shed additional light on the difference between rank-dependent and level-dependent indices by comparing the weighting functions which charac-

| Weighting variable | Version | Rank-dependent | Level-dependent |
|--------------------|---------|----------------|-----------------|
| Socioeconomic status | Basic | \( SR \) | \( SL \) |
| | Extended | \( SR(\nu) \) | \( SL(\alpha) \) |
| Health | Basic | \( HR \) | \( HL \) |
| | Extended | \( HR(\lambda) \) | \( HL(\beta) \) |
terize the two types of indices. We focus here on the case where the weighting variable is socioeconomic status.

The two weighting functions, \( w^r_i(\nu) \) for the rank-dependent indices and \( w^{SL}_i(\alpha) \) for the level-dependent indices, can be visualized in two ways. The first consists of mapping them in function of the ranks \( r_i \), or more precisely, in function of the fractional ranks \( (2r_i - 1)/(2n) \), for successive values of \( i \) (i.e., \( 1/(2n), 3/(2n), \ldots, (2n-1)/(2n) \)). In this way, we can examine how the weights change as one moves up or down the rank order. The second way consists of mapping them in function of the levels \( y_i \), again for successive values of \( i \) (i.e., \( y_1, y_2, \ldots, y_n \)). This allows us to track the evolution of the weights for changing values of the socioeconomic variable.

Figures 1 and 2 illustrate the situation for a typical country. Panel a of Figure 1 represents the rank-dependent weighting function in terms of the fractional ranks, for three different values of the parameter \( \nu \) (\( \nu = 2, 4, 6 \)). For \( \nu = 2 \), the weighting function is linear, as can be seen from expression (5). For higher values of \( \nu \), the curve becomes steeper at the left and flatter at the right, and the point of intersection with the horizontal axis shifts to the left. A rather different picture emerges if one represents the level-dependent weighting function in the same way, again for three different values of the parameter \( \alpha \) (\( \alpha = 0, \frac{1}{2}, 1 \)), as can be seen from Panel a of Figure 2. It is striking, for instance, that for relatively low values of \( \alpha \) the weighting function steeply rises on the right. This occurs when those who are well-off in socioeconomic terms have high levels of the socioeconomic variable. Panel b of Figure 1 represents the rank-dependent weighting function in terms of the levels of the socioeconomic variable. For every value of the parameter \( \nu \) the weighting function has an elongated S-shape. The corresponding representation of the level-dependent weighting function can be found in Panel b of Figure 2. For \( \alpha = 0 \), the weighting function is linear (cf. expression (8)), and for higher values of \( \alpha \), the function becomes steeper at the left and flatter at the right, with the point of intersection with the horizontal axis moving to the left.

[Insert Figures 1 and 2 around here]

Figure 3 highlights the differences between the basic versions of the two types of indices. In this specific example, the level-dependent index gives relatively more weight to the highest ranked individuals, as can be seen from Panel a. Or, to put it differently, the rank-dependent index gives relatively less weight to individuals who attain a high socioeconomic level, as shown

---

4The country we have chosen is the Netherlands. In the following section we provide more details on the socioeconomic variable (equivalent individual income) and the health variable (well-being). Instead of representing each individual separately, we have grouped individuals into percentiles. Each dot represents approximately 1% of the population.
6 An Empirical Illustration

In order to explore whether it makes a difference to use level-dependent rather than rank-dependent indices, we conduct a comparative study on the relation between income and well-being. This study allows us also to find out whether a change in the sensitivity to inequality, either with regard to the socioeconomic distribution or with regard to the health distribution, affects the results. We look at the ranking of a given set of countries according to the different rank-dependent and level-dependent indices we have encountered above. For each of the parameters we take three different values: $\nu, \lambda = 2, 4, 6$, and $\alpha, \beta = 0, \frac{1}{2}, 1$. All in all, we therefore have 12 indices we can use to rank countries.

6.1 The Data

Our data come from the Survey of Health, Ageing and Retirement in Europe (SHARE). We use data from Wave 4, which are derived from surveys taken in 2010-2011 (more details can be found in Malter and Börsch-Supan, 2013). Our socioeconomic variable is based on “total household income” (thinc). The health variable we have selected is the “quality of life and well-being index” (casp), known as CASP-12. This is a variable derived from scores on 12 dimensions of well-being (von dem Knesebeck et al., 2005). Data are available for 16 countries: Austria, Belgium, Czechia, Denmark, Estonia, France, Germany, Hungary, Italy, the Netherlands, Poland, Portugal, Slovenia, Spain, Sweden and Switzerland.

The SHARE data are available at the level of the household (income) and of the individual (well-being). Since income data are often approximate, we have decided to adopt a procedure to smoothen the data. This also helps to avoid that a few observations at the extremes, i.e. very low and very high incomes, distort the results. First, we translated household incomes into equivalent individual incomes using the OECD equivalence scale, which consists of dividing the household income by the square root of the number of household members. Next, we ranked individuals according to their equivalent income and grouped them into percentiles taking into account their sample weights, i.e. we partitioned the population into 100 groups of equal weight. For each percentile we then calculated the mid-interval value.

\[\text{As sample weights, we used the “calibrated cross-sectional weight wave 4” (ciw}_w4)\]
of the equivalent income. This mid-interval income was assigned to every individual of that percentile. In a few countries of our database, a relatively large fraction of the population reported a zero income. For these countries, our procedure would lead to a zero mid-interval income for the first percentile, and maybe also for the second, the third and so on. Suppose that the $k$-th percentile is the first percentile with a non-zero mid-interval income. If $k > 1$, we decided to assign to all individuals of percentiles 1 to $k$ the average income of the percentiles 1 to $k$ taken together. In this way, we always assign a positive income to all individuals. It must be noted, however, that when the fraction of the population with a very low assigned income is considerable, our results may be biased, especially for values of $\alpha$ greater than or equal to 1.

The well-being variable CASP-12 is a bounded variable, which varies between a lower bound of 12 and an upper bound of 48. Its value is calculated as the total score on 12 questions covering different dimensions of well-being: “Respondents were asked, how often they experience certain feelings and situations on a 4-point scale ranging from ‘never’ to ‘often’.” (von dem Knesebeck et al., 2005: 200) The higher the CASP-12 score, the better the quality of life of a person is supposed to be. A value of 39 or higher is considered to indicate a very high quality of life, a value of 37 or 38 a high quality of life, a value of 35 or 36 a moderate quality of life, and a value below 35 a low quality of life. A few key descriptive statistics of our database can be found in Table 1.

6.2 The Results

Table 2 reports the results of the 12 indices. Since our health variable is bounded, we use the absolute version of the $SR$ and $SL$ indices. By contrast, given that our socioeconomic variable is unbounded, we use the relative version of the $HR$ and $HL$ indices. As far as the rank-dependent indices are concerned, these choices are motivated by the arguments advanced because of the irregular nature of the sample weights, it occurred that some groups are slightly larger or slightly smaller than 1%.

6This means that if in a given percentile $y_{\text{min}}$ is the lowest and $y_{\text{max}}$ the highest equivalent individual income, then the mid-interval value is equal to $y_{\text{min}} + (y_{\text{max}} - y_{\text{min}})/2$.

7We applied this procedure to Germany, Hungary, Slovenia and Spain ($k = 2$), and to Italy and Portugal ($k = 4$).

8Given that the individuals in our database do not all have the same sample weight, and that some individuals are tied (i.e., attain the same level of the weighting variable), the formulas of the text need to be modified. The Appendix provides more details on the adjusted formulas.
by Erreygers and Van Ourti (2011). It seems reasonable to apply the same choices to the level-dependent indices.

Table 3 shows the rankings of countries according to all 12 indices. The first observation is that the choice of the index matters. The ranking of countries according to the $SR$ indices is different from that according to the $SL$ indices, and these are in turn different from the rankings according the $HR$ and the $HL$ indices. While the $SR$ indices identify Portugal, Denmark and Czechia as the countries with the lowest degree of pro-rich inequality, and Poland and Germany as the countries with the highest, the $SL$ indices reveal a different pattern: at the lower end we find Denmark, Sweden and the Netherlands, and at the higher end Italy, Estonia and Slovenia. Secondly, the rankings according to the $HR$ and $HL$ indices are more stable than these according to the $SR$ and $SL$ indices. This is probably due to the fact that in general, the health variable is less unequally distributed than the socioeconomic variable. As a result, a change in the distributional parameter $\lambda$ or $\beta$ when well-being is the weighting variable, tends to have less effect than a change in the parameter $\nu$ or $\alpha$ when income is the weighting variable. In what follows, we concentrate on the two types of indices for which income serves as the weighting variable.

Figure 4 illustrates the effect of changes in the distributional parameter on the ranking of countries (In every column the country with the lowest degree of pro-rich inequality is on top, and the country with the highest at the bottom.) Panel a shows the effect of a change in $\nu$ on the ranking of countries according to the rank-dependent indices. As $\nu$ increases, countries like Austria, France, Sweden and Switzerland drop down, while Estonia and Spain make gains. Panel b shows the effect of a change in $\alpha$. Countries like Austria, Germany, Poland and Switzerland see their position in the ranking according to the level-dependent indices worsen, while Czechia and Portugal move up. On the whole, one might say that the rankings are different, that they are influenced by changes in the distributional parameter, but that the effects are limited. A look at the correlation coefficients (Table 4, Pearson’s or Spearman’s rank correlations) reveals that the three $SR$ indices are strongly positively correlated among themselves, especially $SR(2)$ and $SR(4)$, and $SR(4)$ and $SR(6)$. The same seems to hold for the three $SL$ indices, although to a lesser extent for $SL(1)$.

As pointed out above, for some countries the results may be biased for values of $\alpha \geq 1$. The results for $\alpha = 1$ must therefore be treated with caution.

This may be due to the presence of a significant amount of individuals with very low
at the 5% level, namely those between $SR(6)$ and $SL(0)$, between $SR(6)$ and $SL(0.5)$, and between $SR(4)$ and $SL(0)$.

Perhaps the best way of comparing the rank-dependent and the level-dependent indices is to focus on the two basic versions of the indices. The two scatter diagrams of Figure 5 visualize the difference between $SR$ and $SL$, both in terms of the rankings of countries (Panel a) and in terms of the values of the indices (Panel b). Since the two indices use the same information but process it differently, the results are similar but by no means identical. The two rankings are positively correlated, and so are the two values, but in both cases the correlation is far from perfect. For the rankings, the coefficient of correlation $r$ is equal to 0.6882, and for the values, it is equal to 0.5980. Therefore, it does make a difference which index you use. It may very well happen that if one compares the situation in a given country at two different moments of time, one of the indices measures an increase and the other a decrease.\footnote{Since our empirical application is limited to one year only, it cannot be used to illustrate this point.}

\[\text{Insert Figure 5 around here}\]

\section{Discussion}

\subsection{Theoretical Issues}

Above we have argued that, in our view, there are good reasons to move away from the rank-dependent indices which dominate the literature on the measurement of socioeconomic inequality of health. The level-dependent indices we propose as alternatives are, however, not the only ones which can be thought of. To begin with, surely there exist other ways of defining level-dependent weights than the two-part procedure advocated here. Our level-dependent weights consist of a ‘positional’ part which determines the position of an individual relative to others, and an ‘inequality’ part which determines the sum of the absolute values of all the weights. Even if one agrees with that procedure, alternative roads can be followed. In our construction, the positional part is either directly proportional to the level of the weighting variable, or proportional to a transformation of this level. One may think of other forms of transformation than the one we have introduced above, i.e. based upon the isoelastic function. Likewise, the specific inequality measure we have used in our formulas, i.e. the relative mean deviation, is by no means

\section*{22}
the only one available. Alternative measures such as the Gini coefficient or the Theil index could be used just as well. The advantage of the relative mean deviation, however, is that it allows us to establish a direct connection between the level-dependent bivariate inequality measure and the standard statistical concept of covariance.

To sum up, what we propose here is one way of defining a family of level-dependent bivariate inequality measures, which offers some possibilities to take into account distributional sensitivity. A rigorous axiomatic approach might lead to the identification of additional or modified indices.

7.2 Empirical Issues

As far as our empirical application is concerned, the importance of it should not be exaggerated. We have conceived it as an illustration of the potential differences between rank-dependent and level-dependent bivariate inequality measures, not as an in-depth study of the relationship between income and well-being.

The data we have used are far from perfect. Perhaps the most important criticism which can be made of our study is that we treat the CASP-12 variable as a ratio-scale measure of well-being. It would be entirely legitimate to challenge this assumption. If it were a cardinal instead of a ratio-scale variable, it would be inappropriate to use the relative mean deviation as a measure of the inequality in well-being, and therefore to define weights in terms of this variable. This implies that the $HL$ indices, which are based on weights dependent on the levels of the well-being variable, would be unreliable. For this reason, we have chosen to focus on the $SL$ indices, with weights dependent on the levels of income, in our presentation of the results. With regard to the $SL$ indices, we have already pointed out that the occurrence of zero or very low incomes poses a problem for values of the distributional sensitivity parameter greater than or equal to 1.

8 Conclusion

When it comes to the measurement of bivariate inequality, and in particular of the socioeconomic inequality of health, we believe it is time to move from rank-dependent to level-dependent indicators. Rank-dependent indicators ignore quite a lot of valuable information on one of the two dimensions which are being considered. Level-dependent indicators, by contrast, do take that information into account. Moreover, just as rank-dependent indicators, level-dependent indicators can incorporate distributional sensitivity.
But does this matter? Yes, it does: as we have shown by means of an empirical example, the two types of indicators may yield different results. When comparisons are made between countries or over time, rank-dependent and level-dependent indicators do not necessarily produce the same outcome. Although rank-dependent indicators have been used for a long time now and remain popular in empirical research on the socioeconomic inequality of health, we are convinced that level-dependent indicators are to be preferred.

Appendix: Sample Weights and Ties

This appendix describes how the weights $w_i$ must be defined when working with datasets in which not all individuals have the same sample weight, and when there are ties between individuals with respect to the weighting variable. We focus here on the case in which the socioeconomic variable (income, for simplicity) serves as the weighting variable. It is straightforward to extend the formulas to the case in which the health variable serves as the weighting variable.

We assume that our dataset consists of individuals $1, 2, \ldots, n$, ranked according to their individual income, i.e. $y_1 \leq y_2 \leq \ldots \leq y_n$. Let $s_i$ be the sample weight of individual $i$. We assume that $\sum_{i=1}^{n} s_i = 1$.

If there are ties, we define $K$ groups of individuals $G_1, G_2, \ldots, G_K$ such that $y_{G_1} < y_{G_2} < \ldots < y_{G_K}$. The sample weight of group $G_J$ is $s_{G_J} = \sum_{i \in G_J} s_i$. The cumulative sample weight of group $G_J$ is defined by the recursive formula $c_{G_J} = c_{G_{J-1}} + s_{G_J}$, where we take $c_{G_0} = 0$.

Rank-Dependent Weights

Let individual $i$ belong to group $G_J$. Then for $\nu > 0, \nu \neq 1$ the weight of this individual is equal to:

$$w_i^{SR}(\nu) = \left(\frac{1}{\nu - 1}\right) n \left( s_i + \left(\frac{s_i}{s_{G_J}}\right) \left[ (1 - c_{G_J})^\nu - (1 - c_{G_{J-1}})^\nu \right] \right) \quad (A.1)$$

and for $\nu = 1$:

$$w_i^{SR}(\nu) = \left(\frac{s_i}{s_{G_J}}\right) n \left[ (1 - c_{G_J}) \ln (1 - c_{G_J}) - (1 - c_{G_{J-1}}) \ln (1 - c_{G_{J-1}}) \right] \quad (A.2)$$

Level-Dependent Weights

When working with level-dependent weights, there is no need to consider group weights. We do have to modify the definitions of the mean income
\( \mu_{y(\alpha)} \) and of the absolute mean deviation \( \mu_{|d(\alpha)|} \). The mean weighted income is now \( \mu_{y(\alpha)} = \sum_{i=1}^{n} s_i y_i(\alpha) \), and the absolute mean weighted deviation \( \mu_{|d(\alpha)|} = \sum_{i=1}^{n} s_i |y_i(\alpha) - \mu_{y(\alpha)}| \). Taking \( \alpha = 0 \), we obtain \( \mu_y = \sum_{i=1}^{n} s_i y_i \) and \( \mu_{|d|} = \sum_{i=1}^{n} s_i |y_i - \mu_y| \). Using these definitions, the weight of individual \( i \) is equal to:

\[
  w^S_{iL}(\alpha) = \frac{ns_i [y_i(\alpha) - \mu_{y(\alpha)}]}{\sum_{j=1}^{n} s_j [y_j(\alpha) - \mu_{y(\alpha)}]} \cdot \frac{\sum_{j=1}^{n} s_j |y_j - \mu_y|}{\mu_y}
\]  

(A.3)

References

[1] Adler, N.E., T. Boyce, M.A. Chesney, S. Cohen, S. Folkman, R.L. Kahn and S.L. Syme (1994), “Socioeconomic status and health. The challenge of the gradient”, American Psychologist, 49: 15-24.

[2] Bidard, C. (2009), “An axiomatic approach to concentration indices”, EconomiX, Université Paris Ouest Nanterre La Défense, mimeo.

[3] Donaldson, D. and J.A. Weymark (1980), “A single-parameter generalization of the Gini indices of inequality”, Journal of Economic Theory, 22: 67-86.

[4] Erreygers, G. (2009), “Correcting the concentration index”, Journal of Health Economics, 28: 504-515.

[5] Erreygers, G. (2013), “A dual Atkinson measure of socioeconomic inequality of health”, Health Economics, 22: 466-479.

[6] Erreygers, G., P. Clarke and T. Van Ourti (2012), “Mirror, mirror, on the wall, Who in this land is fairest of all?’ – Distributional sensitivity in the measurement of socioeconomic inequality of health”, Journal of Health Economics, 31: 257-270.

[7] Erreygers, G. and T. Van Ourti (2011), “Measuring socioeconomic inequality in health, health care, and health financing by means of rank-dependent indices: A recipe for good practice”, Journal of Health Economics, 30: 685-694.

[8] Kakwani, N.C. (1980), Income Inequality and Poverty. Methods of Estimation and Policy Applications, Oxford University Press, Oxford.

[9] Lambert, P.J. and G. Lanza (2006), “The effect on inequality of changing one or two incomes”, Journal of Economic Inequality, 4: 253-277.
[10] Malter, F. and A. Börsch-Supan (Eds.) (2013), SHARE Wave 4: Innovations & Methodology, Munich, MEA, Max Planck Institute for Social Law and Social Policy.

[11] Marmot, M. et al. (2010), Fair Society, Healthy Lives: The Marmot Review, London: University College London.

[12] Mehran, F. (1976), “Linear measures of income inequality”, Econometrica, 44: 805-809.

[13] O’Donnell, O., E. Van Doorslaer and T. Van Ours (2015), “Health and inequality”, in: A.B. Atkinson and F.J. Bourguignon (Eds.), Handbook of Income Distribution, Amsterdam: Elsevier, Volume 2, Part B, chapter 18, pp. 1419-1533.

[14] Pereira, J.A. (1998), “Inequality in infant mortality in Portugal, 1971-1991”, in: Zweifel, P. (Ed.), Health, the Medical Profession, and Regulation (Developments in Health Economics and Public Policy, vol. 6). Kluwer, Boston/Dordrecht/London, pp. 75-93.

[15] Van Ours T., G. Erreygers and P. Clarke (2014), “Measuring equality and equity in health and health care”, in: A.J. Culyer (Ed.), Encyclopedia of Health Economics, San Diego: Elsevier, Vol. 2, pp. 234-239.

[16] von dem Knesebeck, O., M. Hyde, P. Higgs, A. Kupfer and J. Siegrist, “Quality of life and well-being”, in: A. Börsch-Supan (Coordinator), A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist and G. Weber (Eds.), Health, Ageing and Retirement in Europe. First Results from the Survey of Health, Ageing and Retirement in Europe, Mannheim, Mannheim Research Institute for the Economics of Aging (MEA), pp. 109-203.

[17] Wagstaff, A. (2002), “Inequality aversion, health inequalities, and health achievement”, Journal of Health Economics, 21: 627-641.

[18] Wagstaff, A., P. Paci and E. Van Doorslaer (1991), “On the measurement of inequalities in health”, Social Science and Medicine, 33: 545–57.

[19] Walzer, M. (1983), Spheres of Justice. A Defense of Pluralism and Equality, New York: Basic Books.

[20] Yitzhaki, S. (1983), “On an extension of the Gini inequality index”, International Economic Review, 24: 617-628.
Table 1: Descriptive statistics of the variables under study for the 16 countries.

|          | Equivalent income | Health |          |
|----------|-------------------|--------|----------|
|          | \( n \) | Mean | Median | RMD* | Mean | Median | RMD* |
| Austria  | 4807 | 24054 | 18809 | 0.5488 | 39.60 | 41 | 0.1206 |
| Belgium  | 4787 | 49319 | 24903 | 0.8441 | 37.22 | 38 | 0.1343 |
| Czechia  | 5497 | 7641  | 6355  | 0.5439 | 34.79 | 35 | 0.1350 |
| Denmark  | 2110 | 32997 | 29258 | 0.4172 | 40.72 | 41 | 0.0902 |
| Estonia  | 5997 | 5919  | 4384  | 0.5978 | 35.12 | 36 | 0.1579 |
| France   | 5011 | 26257 | 29240 | 0.5165 | 37.96 | 39 | 0.1254 |
| Germany  | 1469 | 22489 | 17647 | 0.5564 | 38.22 | 39 | 0.1232 |
| Hungary  | 2901 | 5261  | 4307  | 0.5308 | 34.00 | 34 | 0.1650 |
| Italy    | 3365 | 17455 | 12407 | 0.6451 | 33.70 | 34 | 0.1564 |
| Netherlands | 2575 | 30248 | 24272 | 0.5224 | 40.67 | 41 | 0.0944 |
| Poland   | 1629 | 4279  | 3646  | 0.5031 | 35.56 | 36 | 0.1548 |
| Portugal | 1887 | 11163 | 5409  | 0.9698 | 31.87 | 32 | 0.1171 |
| Slovenia | 2574 | 18890 | 9890  | 0.8596 | 39.20 | 40 | 0.1217 |
| Spain    | 3252 | 13417 | 10104 | 0.6152 | 35.81 | 36 | 0.1481 |
| Sweden   | 1860 | 31016 | 26497 | 0.4769 | 38.79 | 39 | 0.1054 |
| Switzerland | 3536 | 70520 | 51656 | 0.6191 | 40.62 | 42 | 0.0967 |

*Relative Mean Deviation
Table 2: Absolute $SR$ and $SL$ indices and relative $HR$ and $HL$ indices for the 16 countries.

| Country     | $SR(\nu)$   | $SL(\alpha)$ | $HR(\lambda)$ | $HL(\beta)$ |
|-------------|-------------|---------------|----------------|-------------|
|             | $\nu = 2$   | $\nu = 4$     | $\nu = 6$     | $\alpha = 0$ | $\alpha = 0.5$ | $\alpha = 1$ | $\lambda = 2$ | $\lambda = 4$ | $\lambda = 6$ | $\beta = 0$ | $\beta = 0.5$ | $\beta = 1$ |
| Austria     | 0.7365      | 0.4685        | 0.3369        | 0.6752       | 0.7829     | 0.8492   | 0.0639       | 0.0422       | 0.0319   | 0.0171       | 0.0173       | 0.0176       |
| Belgium     | 0.6651      | 0.4125        | 0.2808        | 0.9930       | 1.0287     | 1.0816   | 0.0984       | 0.0575       | 0.0406   | 0.0267       | 0.0266       | 0.0263       |
| Czechia     | 0.4981      | 0.2357        | 0.1469        | 0.7689       | 0.6566     | 0.4165   | 0.0769       | 0.0479       | 0.0350   | 0.0221       | 0.0223       | 0.0226       |
| Denmark     | 0.4191      | 0.2580        | 0.1909        | 0.3913       | 0.3876     | 0.3830   | 0.0430       | 0.0341       | 0.0275   | 0.0096       | 0.0099       | 0.0102       |
| Estonia     | 1.0106      | 0.4995        | 0.3076        | 1.5543       | 1.4020     | 1.1870   | 0.1340       | 0.0776       | 0.0548   | 0.0443       | 0.0441       | 0.0439       |
| France      | 0.7402      | 0.4798        | 0.3512        | 0.6993       | 0.7717     | 0.8358   | 0.0683       | 0.0447       | 0.0325   | 0.0184       | 0.0187       | 0.0190       |
| Germany     | 1.0125      | 0.6871        | 0.5309        | 1.1255       | 1.2585     | 1.5130   | 0.1103       | 0.0720       | 0.0538   | 0.0295       | 0.0299       | 0.0303       |
| Hungary     | 0.8968      | 0.4949        | 0.3291        | 1.0562       | 1.0384     | 0.9885   | 0.0875       | 0.0530       | 0.0401   | 0.0311       | 0.0313       | 0.0317       |
| Italy       | 0.9558      | 0.5014        | 0.3297        | 1.7387       | 1.5127     | 1.1600   | 0.1585       | 0.0854       | 0.0578   | 0.0516       | 0.0508       | 0.0500       |
| Netherlands | 0.5113      | 0.3189        | 0.2374        | 0.6054       | 0.6028     | 0.5882   | 0.0701       | 0.0480       | 0.0361   | 0.0149       | 0.0151       | 0.0154       |
| Poland      | 1.1040      | 0.6108        | 0.4024        | 1.0670       | 1.1070     | 1.1317   | 0.0919       | 0.0596       | 0.0448   | 0.0300       | 0.0304       | 0.0307       |
| Portugal    | 0.2770      | 0.0316        | -0.0291       | 0.9755       | 0.8013     | 0.1443   | 0.1300       | 0.0652       | 0.0387   | 0.0306       | 0.0298       | 0.0289       |
| Slovenia    | 0.6740      | 0.3700        | 0.2367        | 1.4010       | 1.3195     | 0.7843   | 0.1460       | 0.0783       | 0.0542   | 0.0357       | 0.0356       | 0.0353       |
| Spain       | 0.7801      | 0.3826        | 0.2313        | 1.1517       | 1.0469     | 0.7481   | 0.1027       | 0.0666       | 0.0486   | 0.0322       | 0.0324       | 0.0327       |
| Sweden      | 0.6256      | 0.4276        | 0.3156        | 0.5513       | 0.5824     | 0.5869   | 0.0606       | 0.0408       | 0.0314   | 0.0142       | 0.0143       | 0.0144       |
| Switzerland | 0.6527      | 0.4366        | 0.3258        | 0.7903       | 0.8608     | 0.9780   | 0.0838       | 0.0568       | 0.0446   | 0.0195       | 0.0199       | 0.0204       |
Table 3: Rankings based on the absolute $SR$ and $SL$ indices and relative $HR$ and $HL$ indices for the 16 countries.

|                | $SR(\nu)$ | $SL(\alpha)$ | $HR(\lambda)$ | $HL(\beta)$ |
|----------------|------------|---------------|----------------|--------------|
|                | $\nu = 2$  | $\alpha = 0$ | $\lambda = 2$ | $\beta = 0$  |
|                | $\nu = 4$  | $\alpha = 0.5$ | $\lambda = 4$ | $\beta = 0.5$ |
|                | $\nu = 6$  | $\alpha = 1$ | $\lambda = 6$ | $\beta = 1$  |
| Austria        | 9          | 4             | 3              | 4            |
| Belgium        | 7          | 9             | 10             | 8            |
| Czechia        | 3          | 6             | 6              | 7            |
| Denmark        | 2          | 1             | 1              | 1            |
| Estonia        | 14         | 15            | 14             | 15           |
| France         | 10         | 5             | 4              | 5            |
| Germany        | 15         | 12            | 12             | 9            |
| Hungary        | 12         | 10            | 8              | 12           |
| Italy          | 13         | 16            | 16             | 16           |
| Netherlands    | 4          | 3             | 5              | 3            |
| Poland         | 16         | 11            | 9              | 10           |
| Portugal       | 1          | 8             | 13             | 11           |
| Slovenia       | 8          | 14            | 15             | 14           |
| Spain          | 11         | 13            | 11             | 13           |
| Sweden         | 5          | 2             | 2              | 2            |
| Switzerland    | 6          | 7             | 7              | 6            |
Table 4: Correlation matrix of country rankings based on the absolute SR and SL indices.

|       | $SR(2)$ | $SR(4)$ | $SR(6)$ | $SL(0)$ | $SL(0.5)$ | $SL(1)$ |
|-------|---------|---------|---------|---------|-----------|---------|
| $SR(2)$ | 1.0000  |         |         |         |           |         |
| $SR(4)$ | 0.9118  | 1.0000  |         |         |           |         |
| $SR(6)$ | 0.7529  | 0.9147  | 1.0000  |         |           |         |
| $SL(0)$ | 0.6882  | 0.4706$^\dagger$ | 0.1735$^\dagger$ | 1.0000 |           |         |
| $SL(0.5)$ | 0.7647  | 0.5971  | 0.3324$^\dagger$ | 0.9765 | 1.0000    |         |
| $SL(1)$ | 0.8735  | 0.9000  | 0.7500  | 0.6441  | 0.7559    | 1.0000  |

Note: all correlations are significant at the 1% level, except for the ones indicated by a '$^\dagger$' sign which are insignificant at the 5% level.
Figure 1: Rank-dependent weighting functions for the $SR(2)$, $SR(4)$ and $SR(6)$ indices in terms of the fractional ranks (a) and the equivalent incomes (b) for the Netherlands.
Figure 2: Level-dependent weighting functions for the $SL(0)$, $SL(0.5)$ and $SL(1)$ indices in terms of the fractional ranks (a) and the equivalent incomes (b) for the Netherlands.
Figure 3: Overlay rank- and level-dependent weighting functions for the basic $SR$ and $SL$ indices in terms of the fractional ranks (a) and the equivalent incomes (b) for the Netherlands.
Figure 4: Visualization of the effect of changes of $\nu$ ($\nu = 2, 4, 6$) on the $SR$ ranking of countries (a) and of $\alpha$ ($\alpha = 0, 0.5, 1$) on the $SL$ ranking of countries (b).
Figure 5: Scatter plots showing the correlation between the basic $SR$ and $SL$ rankings (a) and values (b) for the 16 countries.
Socioeconomic status and health: a new approach to the measurement of bivariate inequality

Date:
2015

Citation:
ERREYGERS, G; KESSELS, R, Socioeconomic status and health: a new approach to the measurement of bivariate inequality, Faculty of Applied Economics Research Paper, 2015

Persistent Link:
http://hdl.handle.net/11343/234009

File Description:
Published version