Inequality with Ordinal Data

by

Frank A. Cowell
STICERD
London School of Economics
Houghton Street
London, WC2A 2AE, UK
e-mail: f.cowell@lse.ac.uk

and

Emmanuel Flachaire
Aix-Marseille University
AMSE and IUF
Centre de la Vieille Charité
2, rue de la Charité
13236 Marseille cedex 02, France
e-mail: emmanuel.flachaire@univ-amu.fr

October 2014
Abstract

The standard theory of inequality measurement assumes that the equal-is-and is a cardinal quantity with known cardinalisation. However, one often needs to make inequality comparisons where either the cardinalisation is unknown or the underlying data are categorical. We propose a natural way of evaluating an individual’s status in such situations, based on their position in the distribution and develop axiomatically a class of inequality indices, conditional on a reference point. We examine the merits of mean, median and maximum status as reference points. We also show how the approach can be applied to perceived health status and reported happiness.

Keywords: Inequality measures, axiomatic approach

JEL codes: D63

Acknowledgments:

Flachaire acknowledges the support of the Institut Universitaire de France (IUF) and of the A*MIDEX project (ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR). We are grateful to Ramses Abul Naga, Thibault Gajdos, Miguel Sánchez, and seminar participants in Alicante, Bath, Sheffield, Université de Laval, Carleton University, Université de Sherbrooke, Université de Montréal and Université Paris 1 Panthéon-Sorbonne for helpful comments.
1 Introduction

It is very common to find problems of inequality comparison where conventional inequality-measurement tools just will not work. A principal cause of this difficulty lies in the nature of the thing that is being studied. By contrast to the inequality analysis of income or wealth distributions where the underlying variable is measurable and interpersonally comparable, “ordinal data” may be measurable but with an unknown scale of measurement, or the underlying concept may not be measurable at all. The term “ordinal data” covers such things as access to amenities, educational achievement, happiness, health, each of which has a sub-literature on inequality in its own right.\footnote{On health status, see for example the discussion by Van Doorslaer and Jones (2003) of the McMaster Health Utility Index (page 65). On life satisfaction and inequality of happiness see for example, Oswald and Wu (2011), Stevenson and Wolfers (2008b), Yang (2008).} This paper provides a new approach to the generic ordinal-data problem.

Why is there a problem with ordinal data? Although we can handle a small number of standard tools of distributional analysis, several key concepts are not well defined. For example the mean will depend on the particular cardinalisation that is used and so there is no meaning to points in a simplex: therefore we cannot implement something like the Principle of Transfers. The literatures on inequality in happiness, health and so on contain a number of work-rounds that address this problem but none of these work-rounds is entirely satisfactory. In some cases first-order dominance criteria have been applied and quantiles have been used to characterise inequality comparisons. But difficulties can arise even with these methods.\footnote{See, for example, Abul Naga and Yalcin (2010), Allison and Foster (2004).}

However, in this paper we show that, as long as the underlying data can be ordered, it is possible to construct a fully-fledged approach to inequality measurement. In section 2 we discuss the basic building blocks of inequality analysis, section 3 develops our formal approach to the problem of ordinal data and section 4 discusses the properties of the class of inequality measures that emerges from our analysis. Section 5 investigates the empirical properties of the class of inequality measures and discusses two important applications; section 6 concludes.
2 Inequality analysis: basics

There are three basic ingredients of the inequality-measurement problem: (1) the definition of the equalisand, the thing that is unequally distributed; (2) the definition of the receiving unit, the basic component of the population amongst which there is an unequal distribution; (3) a method of aggregation.

In what follows we will assume that there is agreement on ingredient (2): let us say that inequality is to be seen as an unequal distribution among persons. However there are several possibilities for ingredient (1) – such as income, utility, status – that differ in terms of their informational requirements and that require different methods of aggregation.

2.1 Cardinal data

Let us begin with the most common way of representing the inequality-measurement problem. This takes the first ingredient of the measurement problem as a cardinal entity, such as income, wealth, or expenditure. In adopting this approach we are assuming the equalisand to be both measurable and inter-personally comparable (Cowell 2011).

Income

For a given finite population an income distribution is simply a vector of real numbers. Because the components of the vector have cardinal significance, calculations of the total and the mean are well defined and it becomes possible to apply some second-order dominance criteria. But the income-inequality measurement problem involves a tighter restriction than just a requirement that the equalisand be a cardinal entity. Typically we restrict the analysis to a specific scale and origin: having done so we may then make comparisons of distributions that are expressed as shares in total income, such as Lorenz curves.

Utility

Suppose that we are concerned with the problem of measuring the inequality of utility rather than inequality of something like income. In some cases it is reasonable to assume a one-to-one relationship between utility and income.
In other words there is a relationship

\[ u = U(x) \] (1)

where \( x \) is an “income” variable that has the specific cardinal qualities just described and \( U \) is a known monotonic function. In this case it may be possible to infer the shape of the function \( U \) from, for example, people’s attitude to risk.

If utility were an affine transformation of income, then the simple second-order dominance relationship among distributions (“Generalised-Lorenz” comparisons) is preserved as we go from income to utility, but regular Lorenz comparisons and inequality measures are not preserved.\(^3\) But, if there is also agreement on the scale and origin of \( u \) itself, then clearly one could apply the methods suitable for income inequality to \( u \) instead of \( x \).\(^4\)

Otherwise, if a monotonic transformation of \( U \) is regarded as valid as \( U \) itself, then we could adopt an approach whereby the problem of the inequality of \( u \) is seen as equivalent to the problem of the inequality of \( x \). We may imagine using the inverse of (1) to convert \( u \)-values back into \( x \)-values in order to address the problem of utility-inequality.\(^5\)

### 2.2 Ordinal data

We now consider cases where the equalisand is less rich in information than the case cardinal-data cases as described in Section 2.1. Ordinal data can be modelled in several ways, of which we will consider two leading cases.

---

\(^3\) As an example consider the Gini coefficient if utility is subjected to the following transformation: multiply every person’s utility by an amount \( \lambda \) and add an amount \( \delta = \mu(1-\lambda) \) where \( \mu \) is the mean value of pre-transformation utility; if \( \lambda < 1 \) this transformation utility will automatically reduce measured inequality.

\(^4\) If one allows the function \( U \) to incorporate other personal information about needs then the equivalised-income approach can be seen as an example of this.

\(^5\) There is an analogy with the Atkinson (1970) revolution in inequality measurement. Atkinson pointed out that, although the Dalton (1920) insight of using (social) utility was a valuable step in the appropriate direction for introducing social values into inequality measurement, looking at the relative shortfall of average utility of income below the utility of average income was flawed, in that it was dependent on the assumed origin of the utility scale. Atkinson’s solution was to introduce the concept of equivalent income. However Atkinson and Dalton were principally concerned with the aggregation part of the measurement problem: the way in which social values could be introduced into an inequality-evaluation of income distribution, not the inequality-evaluation of a distribution of utilities.
A transformation of “income”

Again let us suppose that a relationship (1) holds but, by contrast to Section 2.1, that the form of $U$ is unknown or that $x$ is a measurable quantity that has no agreed valuation. If the form of $U(\cdot)$ is unknown and cannot be inferred from individual behaviour, then increased dispersion of $x$ leads to increased inequality, but we cannot say by how much. If $U$ differs across individuals we cannot say even that.

The quality of life and happiness are two practical examples of this case.

Categorical data

Alternatively, suppose utility depends on a categorical variable. We need only to make the assumption that the categories can be ordered in a well-defined fashion to provide a simple approach to inequality comparisons.

As an illustrative example of the concept consider the problem of inequality in access to specific public facilities. Assume that there is general agreement that, other things being equal, a person is better off with access to both gas and electricity supplies than with access to electricity only, a person is better off with access to electricity only than gas only and is better off with access to gas only than access to neither utility; we know nothing about how much energy is consumed or about how much could be afforded; so trying to put a dollar equivalent may be inappropriate or meaningless. In Table 1 suppose $n_k$ is the number of persons in category $k \in \{B, E, G, N\}$. If there are 100 people in the population, intuition suggests that Case 1 in Table 1 represents more amenity inequality than does Case 2.

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_k$</td>
<td>$n_k$</td>
</tr>
<tr>
<td>Both Gas and Electricity</td>
<td>25</td>
</tr>
<tr>
<td>Electricity only</td>
<td>25</td>
</tr>
<tr>
<td>Gas only</td>
<td>25</td>
</tr>
<tr>
<td>Neither</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 1: Access to amenities: categorical variable

Inequality of educational attainment or health status can be seen as other practical examples of this type of problem.
Now consider the implications of these two cases for the way in which one approaches the inequality-measurement problem. Although, as we have seen, there is a natural way of dealing with utility-inequality in the cardinal-data setting, the method adopted there is not of immediate help in the measurement of inequality comparisons in the purely ordinal setting. In the face of these difficulties two main approaches have been adopted in the literature.

The first approach is to force a solution on the problem by imputing some artificial index of individual well-being as a function of \( x \) or of the relevant category. Sometimes the imputation is achieved through subjective evaluation by individuals (for example on a Likert scale) and sometimes by official institutions (for example the Quality-Adjusted Life Year or the Human Development Index). Clearly the same procedure can be applied explicitly or implicitly to entities that do not have a natural ordering, such as vectors of attributes or endowments; one uses the utility function to force an ordering of the data.\(^6\) This approach runs into objections such as the arbitrariness of the cardinalisation, the arbitrariness of aggregating apples and oranges, and the arbitrariness of attempting to include measures of dispersion into the index as well. Even if the resulting \( y \) appears reasonable over a wide subset of the possible values of \( x \) we might still be concerned about the way extreme values are represented in the index and their consequences for inequality comparisons.

The second approach, common in the health literature,\(^7\) involves a reworking of traditional inequality-ranking approaches focusing on first-order dominance criteria. The median has been suggested as an equality concept corresponding to the use of the mean in conventional inequality analysis, although it has been noted that comparing distributions with different medians raises special issues (Abul Naga and Yalcin 2010). Clearly such an approach may run into difficulty if quantiles are not well-defined, as may happen in the case of categorical variables.\(^8\)

\(^6\)This is similar to one of the standard theoretical approaches to the measurement of multi-dimensional inequality – one computes the “utility” of factors and then computes inequality of utility where the utility function is an appropriate aggregator (Maasoumi 1986, Tsui 1995).

\(^7\)For the application to health, see Abul Naga and Yalcin (2008), Allison and Foster (2004), Zheng (2011).

\(^8\)For examples of this see footnote 19 below.
2.3 An alternative approach

Since both of the standard approaches to the ordinal-data case are unsatisfactory to some extent it is useful to consider an alternative approach.

In the income-distribution literature it is common to find a person’s location in the distribution used as a concept of status in society: this concept has then been used to develop measures of individual and social deprivation and has sometimes been incorporated into the measurement of inequality. Here it is central to the alternative approach. Denote by $u_i$ person $i$’s utility and let the distribution function of $u$ for a population of size $n$ be $F$. Each person’s status $s_i$ is uniquely defined for a given distribution as $s_i = \psi(u_i, F(\cdot), n)$, where the function $\psi$ is such that status is independent of the cardinalisation of utility. One simple way of specifying status for an individual is the standard definition of position, the proportion of the population that is no better off than oneself; so if the distribution cumulative function for utility in the population is $F$ then we could use $s_i = F(u_i)$ as the status measure.$^9$

Such a status measure is familiar to anyone who has ever taken a GRE or TOEFL test and is similar to that sometimes used to measure opportunity – see de Barros et al. (2008). Working with the status thus implied by utility rather than with utility itself would obviously dispose of the problem of cardinalisation, as illustrated in Figure 1: if $U$ and $V$ are two alternative cardinalisations of the utility of income $x$ ($V$ is a monotonic increasing transformation of $U$), then the two utilities for person 1 under the two cardinalisations $u_1$ and $v_1$ each map into $s_1$ and, for person 2, $u_2$ and $v_2$ each map into $s_2$.

However, for categorical data – where we know only the ordering of the categories $k$ and the number of persons $n_k$ in each category, $k = 1, 2, ..., K$ – there is a stronger reason for taking a status measure of this form. Consider the following “merger principle”:

**Mergers.** If non-empty adjacent categories $k^*$ and $k^* + 1$ are merged then this has no effect on the status of a person who belongs to neither category.

We can easily see that this simple principle induces an additive structure. Suppose there are three categories and that person $i$ belongs to category 1. Now let categories 2 and 3 be merged; in view of the mergers principle $^9$ $F(u_i)$ and $1 - F(u_i)$ correspond to Yitzhaki (1979)’s concepts of “Satisfaction” and “Deprivation.”
person \( i \)'s status is given by \( s_i = f_i(n_1, n_2, n_3) = f_i(n_1, n_2 + n_3, 0) \); if there were \( K \) categories rather than three and categories 2, \ldots, \( K \) were merged then, applying the merger principle iteratively, \( s_i = f_i(n_1, n_2, \ldots, n_K) = f_i\left(n_1, \sum_{\ell=2}^{K} n_\ell, 0, \ldots, 0\right) \). Therefore, if the status function is anonymous, so that every person in a given category has the same status, it is clear that status must take the form

\[
s_i = f\left(\sum_{\ell=1}^{k(i)-1} n_\ell, n_{k(i)}, \sum_{\ell=k(i)+1}^{K} n_\ell\right),
\]

where \( k(i) \) is the category to which person \( i \) belongs. Hence we could take person \( i \)'s status to be “downward-looking”, in other words of the form

\[
\sum_{\ell=1}^{k(i)} n_\ell, \text{ for all } i
\]
or
\[
\sum_{\ell=1}^{k(i)-1} n_{\ell}, \text{ for all } i \text{ such that } k(i) > 1 \text{ and 0 otherwise.} \quad (3)
\]

(my status is determined by all those below me and possibly by my peers, those in the same category as me). We could also consider the “upward-looking” counterparts of (2) and (3):

\[
\sum_{\ell=k(i)}^{K} n_{\ell}, \text{ for all } i, \quad (4)
\]

\[
\sum_{\ell=k(i)+1}^{K} n_{\ell}, \text{ for all } i \text{ such that } k(i) < K \text{ and 0 otherwise.} \quad (5)
\]

Here my status is determined as the absolute numbers of those above me and perhaps also my peers (as in a football league). Other formulations of status may be appropriate: in particular the status measures corresponding to (2)-(5) that have been normalised by dividing through by \( n \). However, we do not need to specify a particular form for our theoretical approach; in section 4 we will provide particular examples based on equations (2)-(5).

3 Inequality measurement: theory

3.1 Approach

Here we offer a new approach to inequality measurement that draws on the themes discussed in section 2, but which does more. It provides a coherent general approach to inequality measurement, within which is nested a method for ordinal data. The same framework of analysis can be used for the well-known problem with cardinal data and the much less well explored case of categorical data. The approach involves two steps.

Step 1: define status. The precise definition of status will depend on the structure of information and may also depend on the purpose of the inequality analysis. In some cases a person’s status is self-defining from the data: for example if we want to focus on wealth inequality then we use a measure of net worth. In some cases status is defined once one is given additional distribution-free information: for example if there are
observations on some variable $x$ and it is known that utility is $\log(x)$. In some cases status requires information dependent on distribution of the underlying data: for example it could be determined as one of the specifications (2)-(5). For the moment we need only assume that an individual’s status is given by

$$s \in S \subseteq \mathbb{R}.$$  

**Step 2: summarise the status distribution.** The distribution of status is given by a vector $s \in S^n$, where $n$ is the size of the population. To make inequality comparisons one needs to aggregate the information provided by different $s$-vectors: to make progress on this step requires both a concept of equality and a way of characterising departures from equality.

Let us briefly consider these requirements for step 2. We define $e \in S$ as an equality-reference point, which could be exogenously given, or could alternatively depend on the status vector

$$e = \eta(s).$$  

(6)

In the conventional income-inequality case (where the summation of incomes is well defined) it makes sense to assume that equality is where everyone receives mean income; the specification (6) is appropriate, where the function $\eta$ is simply the mean. However, for other types of data this specification is inappropriate: for example the mean may not be well-defined if status is given in terms of utility. The specification of $e$ is discussed further in Section 4.2 below. To capture inequality we could define a specific distance function $d : S^2 \rightarrow \mathbb{R}$, where $d(s, e)$ means the distance that a person with status $s$ is from $e$, the reference point,\(^{10}\) and then introduce a number of principles (axioms) to characterise an inequality ordering $\succeq$ and the associated distance concept. However, we do not need to do as much as this.

As we will see in sections 3.2-3.4 we can make progress without specifying an explicit function $d(\cdot)$ *a priori*. By characterising $\succeq$ axiomatically the distance-from-equality concept will emerge.

\(^{10}\)This is analogous to the interpretation of Generalised-Entropy measures of income inequality that can be thought of as average distance from mean income.
3.2 Inequality ordering: basic structure

Consider inequality as a weak ordering $\succeq$ on $S^{n+1}$; denote by $\succ$ the strict relation associated with $\succeq$ and denote by $\sim$ the equivalence relation associated with $\succeq$. For any vector $s$ denote by $s(\varsigma, i)$ the vector formed by replacing the $i$th component of $s$ by $\varsigma \in S$. We first characterise the general structure of the inequality relation using just four axioms:

**Axiom 1 [Continuity]** $\succeq$ is continuous on $S^{n+1}$.

**Axiom 2 [Monotonicity in distance]** If $s, s' \in S^n$ differ only in their $i$th component then (a) if $s'_i \geq e : s_i > s'_i \iff (s, e) \succ (s', e)$; (b) if $s'_i \leq e$ \iff $(s, e) \succ (s', e)$.

**Axiom 3 [Independence]** For $s, s' \in S^n$ , if $(s, e) \sim (s', e)$ and $s_i = s'_i$ for some $i$ then $(s(\varsigma, i), e) \sim (s'(\varsigma, i), e)$ for all $\varsigma \in [s_{i-1}, s_{i+1}] \cap [s'_{i-1}, s'_{i+1}]$.

**Axiom 4 [Anonymity]** For all $s \in S^n$ and permutation matrix $\Pi$, $(\Pi s, e) \sim (s, e)$.

**Theorem 1** Given Axioms 1 to 4 $\succeq$ is representable by the continuous function $I : S^{n+1} \rightarrow \mathbb{R}$ given by

$$I (s; e) = \Phi \left( \sum_{i=1}^{n} d(s_i, e), e \right), \quad (7)$$

where $d : S \rightarrow \mathbb{R}$ is a continuous function that is strictly increasing (decreasing) in its first argument if $s_i > e$ ($s_i < e$).

It is clear that to obtain this first result we need only standard and easily interpretable assumptions. Monotonicity (Axiom 2) gives meaning to the inequality relation: if two distributions differ only in respect of person $i$’s status, then the distribution that registers greater individual distance from equality for $i$ is the distribution that exhibits greater inequality. Independence (Axiom 3) means the following: suppose distributions $s$ and $s'$ are equivalent in terms of inequality and that there is some person $i$ who has

---

11Here we follow the method of Cowell and Ebert (2004) and Ebert and Moyes (2002).

12Proofs of theorems are given in the Appendix.
the same status in $s$ and in $s'$; then the same small change in $i$’s status in both distributions $s$ and $s'$ still leaves $s$ and $s'$ as equivalent in terms of inequality.\footnote{This rules out aggregations of status that incorporate people’s distributional ranks in the aggregation formula; but distributional rank is of course already taken account of in the specification of individual status.} Anonymity (Axiom 4) means that if all information relevant to inequality is embodied in the status measure then permuting the labels on the individuals must leave inequality unchanged.

The resulting Theorem 1 establishes inequality as total “distance” from equality; the function $d$ is continuous and has the property that, for a given $e$, $d(s,e)$ is increasing in status $s$ if $s$ is above the reference point, decreases in $s$ if $s$ is below the reference point; the nature of $d$ will be discussed further in section 4.\footnote{Note that $d$ is not a conventional distance function since the symmetry property is not required.}

### 3.3 Inequality ordering: scale

Of course Theorem 1 gives only a general outline of the type of inequality measures that can be applied to both cardinal and ordinal data. To make progress we next introduce more structure on $d$ and hence on $\succeq$. The key property is that inequality orderings remain unchanged when status is rescaled. Two variants of this property should be considered.

**Axiom 5 [Scale invariance 1]** For all $\lambda \in \mathbb{R}_+$: if $s, s', \lambda s, \lambda s' \in S^n$ and $e, e' \in S$ then $(s, e) \sim (s', e') \Rightarrow (\lambda s, e) \sim (\lambda s', e')$.

**Axiom 6 [Scale invariance 2]** For all $\lambda \in \mathbb{R}_+$: if $s, s', \lambda s, \lambda s' \in S^n$ and $e, e', \lambda e, \lambda e' \in S$ then $(s, e) \sim (s', e') \Rightarrow (\lambda s, \lambda e) \sim (\lambda s', \lambda e')$.

Notice that in Axiom 5 we do not require that the reference point be the same in the two distributions being compared, just that, for given $n$, it is independent of the vector $s$. In Axiom 6 the reference point in each distribution is rescaled along with the status distribution itself. Corresponding to the two versions of scale invariance we have two key results.

**Theorem 2** Given Axioms 1 to 5 $\succeq$ is representable by (7) where the function $d$ takes the form

$$d(s,e) = A(e) s^{\alpha(e)}$$

and $A$ and $\alpha$ are real numbers that may depend on $e$. 

Theorem 3  Given Axioms 1 to 4 and 6 \( \succeq \) is representable by (7) where the function \( d \) takes the form

\[
d(s, e) = e^\beta \phi \left( \frac{s}{e} \right).
\]

(9)

where \( \beta \) is a constant and \( \phi \) is an arbitrary function.

Which form of the scale-invariance property should be introduced depends on the reference point. If \( e \) is independent of \( n \) then clearly Scale invariance 1 (Axiom 5) is appropriate, whereas if \( e \) varies with \( n \) then Scale invariance 2 (Axiom 6) is appropriate. The scale-invariance assumption has an appealing interpretation for both types of data structure

Ordinal data

Both Axiom 5 and Axiom 6 are relevant to the ordinal-data problem, as can be seen by the following two examples, where \( e \) is taken as a person’s status in the case of a perfectly equal distribution:

1. If status is given by the peer-exclusive form (5) then, if everyone is in the same category (perfect equality), each person’s status is zero: \( e = 0 \).

2. If status is given by the peer-inclusive form (4) then, if everyone is in the same category, each person’s status is given by \( e = n \).

Adopting either or both forms of scale-invariance means that, in the case of ordinal data, we can express status in terms of proportions of the population (where expressions (2)-(5) are normalised by dividing through by \( n \)) and that inequality is to be expressed in terms of the dispersion of those proportions. The property is equivalent to the Dalton principle of population.\(^{15}\)

Cardinal data

In the case where status is cardinal and is measured by income, scale invariance implies that inequality can be measured in terms of income shares.\(^{16}\)

\(^{15}\)See also Yaari (1988) for a similar assumption.

\(^{16}\)See, for example, Cowell and Kuga (1981a, 1981b).
3.4 Inequality ordering: Basic form

Theorems 2 and 3 give us a complete ordering of distributions by inequality and the essentials for family of inequality measures.

To see how this works consider the joint application of (8) and (9) in two cases corresponding to different values of \( e \). (1) If \( e \neq 0 \) then, from (8) and (9) we have

\[
d(s,e) = A(e) s^{\alpha(e)} = e^\beta \phi \left( \frac{s}{e} \right)
\]

and, putting \( s = e \), we have

\[
A(e) e^{\alpha(e)} = ce^\beta,
\]

where \( c := \phi(1) \); so, substituting back into (10), we have

\[
\phi(z) = cz^{\alpha(e)}, \quad \text{where } z := s/e.
\]

Since \( s \) is arbitrary, \( z \) is arbitrary and clearly \( \alpha(e) \) must be a constant, independent of \( e \), so that \( \phi(z) = cz^\alpha \). (2) If \( e = 0 \) then Theorem 2 implies that \( d(s,e) = cs^\alpha \), where \( c := A(0) \) and \( \alpha := \alpha(0) \). Therefore in general we have

\[
d(s,e) = cs^\alpha e^{\beta - \alpha},
\]

with the additional restriction \( \beta \geq \alpha \) if \( e \) can take the value 0.

Plugging (11) into (7) gives the general functional form representing the inequality ordering. Given the general transformation \( \Phi \) included in (7) it is possible to normalise so as to choose the scale and origin of \( I \) appropriately. Accordingly we have that, given Axioms 1 to 6, \( \succeq \) is representable as

\[
I_\alpha(s;e) := \frac{1}{\alpha[\alpha - 1]} \left[ \frac{1}{n} \sum_{i=1}^{n} s_i^\alpha - e^\alpha \right],
\]

where \( \alpha \in \mathbb{R} \), or by some strictly increasing transformation of (12) involving \( e \). Clearly this normalisation ensures that \( I_\alpha(s;e) \) takes the value 0 if \( s_i = e \) for all \( i \). Other properties are examined in section 4.

4 Inequality measures

As we have seen, the conventional approach to inequality measurement only works within a narrowly defined information structure. In our alternative
Table 2: Distributional comparisons for a categorical variable

<table>
<thead>
<tr>
<th>Case 0</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_k$</td>
<td>$s_i$</td>
<td>$s'_i$</td>
<td>$n_k$</td>
</tr>
<tr>
<td>B</td>
<td>$0$</td>
<td>$25$</td>
<td>$1$</td>
</tr>
<tr>
<td>E</td>
<td>$50$</td>
<td>$1$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>G</td>
<td>$25$</td>
<td>$1/2$</td>
<td>$3/4$</td>
</tr>
<tr>
<td>N</td>
<td>$25$</td>
<td>$1/4$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

$\mu(s) = \frac{11}{16}$  $\frac{5}{8}$  $\frac{3}{4}$  $\frac{11}{16}$

approach how do we proceed to get a usable inequality index? Equation (12) provides us with the structure for inequality analysis; to get from here to an operational inequality measure with ordinal data requires some further steps. We need to check that measures of the form $\Phi(I(s; e), e)$ have sensible properties when taken applied to ordinal data, we need to consider how the reference point for inequality comparisons is to be determined, and we need to clarify the way in which different members of the class of indices of the form (12) behave.

4.1 The transfer principle?

In standard approaches to inequality measurement the transfer principle plays a central role. So it is reassuring to note that if status is cardinal and is measured by income or wealth the transfer principle is satisfied by (12). This is because $I_{\alpha}(s; e)$ is increasing and convex in each $s_i$ so that an infinitesimal transfer of $\delta$ from $j$ to $i$ must reduce $I_{\alpha}(s; e)$ if $s_i < s_j$.

The transfer principle is inappropriate in the case of ordinal data because there is no natural “compensation” to consider, but one might wonder whether a modified form of this principle is valid. Consider a simple example using the access-to-utilities story.

Table 2 represents four distributional scenarios based on the example in Table 1. As before $n_k$ is the number of persons in category $k \in \{B, E, G, N\}$ and in each scenario there are 100 persons in the population. We consider the two versions of (normalised) peer-inclusive status introduced in subsection
2.3:

\[ s_i = \frac{1}{n} \sum_{\ell=1}^{k(i)} n_\ell \text{ (downward looking)}, \quad (13) \]

\[ s'_i = \frac{1}{n} \sum_{\ell=k(i)}^{K} n_\ell \text{ (upward looking)}, \quad (14) \]

- see equations (2) and (4). The last row in the table contains the value of mean status,

\[ \mu(s) = \mu(s') = \frac{1}{n} \sum_{i=1}^{n} s_i = \frac{1}{n} \sum_{i=1}^{n} s'_i, \quad (15) \]

If the relevant scenario changes from Case 0 to Case 1 then 25 people are promoted from category E to category B and, by the principle of monotonicity, if \( e \) were a constant equal to any of the values taken by \( \mu(s) \) (namely \( 5/8, 11/16 \) or \( 3/4 \)) then inequality would increase; if instead the relevant scenario changes from Case 0 to Case 2 then 25 people are promoted from category N to category G and, by the same principle, inequality would decrease.

Now suppose we try to invoke a principle that appears to be related to the transfer principle. If \( e \) is a constant, independent of \( s \) or if \( e \) depends only on \( \mu(s) \) then, the following is true for all values of \( \alpha \): if there is a change in the underlying distribution such that \( i \)'s status increases by \( \delta > 0 \) and \( j \)'s status decreases by \( \delta \) (where \( s_i < s_j \) and \( s_i + \delta < s_j - \delta \), then, from (12) inequality is reduced. However, this "transfer property" is not particularly attractive, as we can see from Table 2. If the relevant scenario changes from Case 0 to Case 3 this exactly fits the balanced change-of-status story: 25 persons promoted from N in Case 0 to G in Case 3 experience an increase in status of \( 1/4 \); because 25 persons are promoted from E to B this reduces the status of those left in category E by \( 1/4 \). But the change from Case 0 to Case 3 is more appropriately seen as a combination of an inequality-increasing change (Case 0 to Case 1) and inequality-decreasing change (Case 2 to Case 3).\(^{17}\)

What is important is the individual move towards or away from the reference point.

\(^{17}\)Equivalently we could see it as a combination of the inequality-decreasing change Case 0 to Case 2 and the inequality-increasing change Case 2 to Case 3.
4.2 The reference point

The importance of specifying a reference point is evident from the example just given. The behaviour of $I_\alpha$ in (12) may be sensitive to the choice of $e$. What values could or should $e$ take? In the case of cardinal data the choice is obvious (see option III below). In the case of ordinal data we consider four possibilities: two where reference point $e$ is exogenous and two where it depends on $s$ as in equation (6) so that (12) becomes

\[ I_\alpha (s; \eta (s)) = \frac{1}{\alpha (\alpha - 1)} \left[ \frac{1}{n} \sum_{i=1}^{n} s_i^\alpha - \eta (s)^\alpha \right]. \tag{16} \]

In discussing each of these possibilities we take the status of individual $i$ to be $i$’s position in the distribution given by (13), (14) or their peer-exclusive counterparts.

I maximum status

Consider what happens in the situation of perfect equality where the distance from the reference point must be zero for everyone. If status is is peer-inclusive as in (13) or (14), there is only one situation in which we can have $s_i = e, i = 1, ..., n$ in (16): this is where $e = 1$, the maximum possible value of $s$. However, although it is clear that $I_\alpha (1; 1) = 0$ for all $\alpha$ (the measure is always zero for an equal distribution) there is no guarantee that $I_\alpha$ will be non-negative for every value of $\alpha$: this problem is discussed further in section 4.3.

II minimum status

In the corresponding peer-exclusive version of status one modifies (13) and (14) by excluding category $k (i)$ from the summation. Again there is only situation where $s_i = e, i = 1, ..., n$: this time it is where $e = 0$, the minimum value of $s$. Immediately we see from (16) that $I_\alpha$ will only be defined for $\alpha \geq 0$.

III mean status

\[^{18}I_\alpha (s; \eta (s)) \text{ is invariant under replications of the population and } I_\alpha (e; e) = 0. \text{ Other properties will depend on the specification of } \eta (\cdot).\]
In the conventional approach to inequality measurement the reference point is taken to be the mean value of the quantities to which the inequality measure is being applied – mean income, mean wealth or whatever. Although one does not suppose that this mean value is what individuals would actually receive if a policy were implemented to enforce equality, it is still a useful and informative reference point. Taking the lead from this we might consider
\[ e = \eta(s) = \mu(s) \]
given in (15). If we use this specification in (16) then it is easy to show that \( I_\alpha(s; \mu(s)) \geq 0 \) with equality if and only if \( s_i = \mu(s) \), for all \( i \). It is also the case that \( I_\alpha(s; \mu(s)) \) is continuous in \( \alpha \). Note that, by contrast to standard inequality analysis where one conventionally assumes a fixed total of income or wealth, in the case of categorical data \( \mu(s) \) cannot be an exogenously given constant (see, for example, Table 2).

IV median status

Consider the median, \( e = \eta(s) = \text{med}(s) \), defined as \( e \in S \) such that \( \#(s_i \leq e) \geq \frac{n}{2} \) and \( \#(s_i \geq e) \geq \frac{n}{2} \), as a possible reference point. We immediately find a fundamental problem for the kind of problem considered here: in the case of categorical data the median is not well-defined. For example in Table 2 the median could be any value in an interval \( M(s) \) where \( M(s) = [1/2, 1] \) in cases 0 and 2 and \( M(s) = [1/2, 3/4] \) in cases 1 and 3. Even if we resolve this problem by picking one specific value \( e \in M(s) \) as the reference point, for example the lower bound of the interval \( M(s) \) it is not clear that this provides an appropriate reference point with categorical data. Furthermore, there is nothing in the formula (16) that prevents the index taking a negative value.

Which of these these alternative concepts of the reference point should be used? Let us compare them using the amenity-inequality example and peer-inclusive downward-looking status. The first three rows of Table 3 shows the

---

19Two examples illustrate the kind of problem that can arise. (1) With three ordered categories and the same proportion of individuals in each category, the median is ambiguous. The status vector is \( s = (1/3, 2/3, 1) \). The conventional definition of the median gives \( \text{med}(s) = m := 2/3 \). But, nevertheless we can see that \( 2/3 \) of the population has a status less or equal to \( m \) and \( 2/3 \) of the population has a status greater than or equal to \( m \). (2) With two ordered categories (where B is better than A) and a thousand persons consider the following three distributions: (i) \( n_A = 500, n_B = 500 \); (ii) \( n_A = 499, n_B = 501 \); (iii) \( n_A = 999, n_B = 1 \); the corresponding status vectors are (i) \( s = (0.5, 1) \), \( \text{med}(s) = 0.5 \); (ii) \( s = (0.499, 1), \text{med}(s) = 1 \) (iii) \( s = (0.999, 1), \text{med}(s) = 0.999 \). Distributions (i) and (ii) have very different medians, but distributions (ii) and (iii) have almost the same median!
Table 3: Inequality comparisons for a categorical variable: Downward-looking status

<table>
<thead>
<tr>
<th></th>
<th>Case 0</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu(s)$</td>
<td>$\frac{11}{16}$</td>
<td>$\frac{5}{8}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{11}{16}$</td>
</tr>
<tr>
<td>med$_1(s)$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{5}{8}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{5}{8}$</td>
</tr>
<tr>
<td>med$_2(s)$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>$I_0(s; \mu(s))$</td>
<td>0.1451</td>
<td>0.1217</td>
<td>0.0588</td>
<td>0.0438</td>
</tr>
<tr>
<td>$I_0(s; \text{med}_1(s))$</td>
<td>0.2321</td>
<td>0.1217</td>
<td>0.0588</td>
<td>-0.0515</td>
</tr>
<tr>
<td>$I_0(s; \text{med}_2(s))$</td>
<td>-0.1732</td>
<td>-0.1013</td>
<td>-0.3465</td>
<td>-0.2746</td>
</tr>
<tr>
<td>$I_0(s; 1)$</td>
<td>0.5198</td>
<td>0.5917</td>
<td>0.3465</td>
<td>0.4184</td>
</tr>
</tbody>
</table>

Values of a number of reference points $e$, the mean $\mu(s)$ and two possible interpretations of the median: med$_1(s)$ is the midpoint of the interval $M(s)$ and med$_2(s)$ is the lower bound of $M(s)$. The four columns correspond to the four cases in Table 2; notice that $\mu(s)$ and med$_1(s)$ differ from one distribution to another. Rows 4-7 of Table 3 give the values of the index $I_0(s; e)$ (see (17) below) for each of these three endogenous specifications of $e$ and for the case $e = 1$.

The specifications $I_0(s; \mu(s))$ and $I_0(s; \text{med}_1(s))$ appear to produce counterintuitive results: inequality decreases when one person is promoted from E to B (Case 0 to Case 1, or Case 2 to Case 3). We can understand why this happens when we see that the movement of this person changes both $\mu(s)$ and med$_1(s)$ reference points. By contrast, if we use the med$_2(s)$ specification, the reference point does not change and the inequality changes are in the direction that accords with intuition; the major problem, of course, is that $I_0(s; \text{med}_2(s))$ is negative for all the cases in this example!

Only one specification of the reference point appears to produce an inequality measure that behaves consistently with what one might expect: it is the last row with $e = 1$. Here we find that $I_0(s; 1)$ increases when one individual moves away from the (fixed) reference point (Case 0 to Case 1, Case 3 to Case 1) and $I_0(s; 1)$ decreases when one individual moves toward the reference point (Case 0 to Case 2, Case 3 to Case 2). Moreover measured inequality is always positive if $s \neq 1$.

It seems that the use of either definition of the median produces unsatisfactory results – some or all of the inequality values are negative in our simple example. Use of mean status seems to produce strange results when the mean changes significantly as people’s position changes. This appears to
leave only maximum status as a candidate reference point. To get further
insight on this let us consider examine the way the class (16) behaves for
different values of $\alpha$.

4.3 The sensitivity parameter $\alpha$

The parameter $\alpha$ in the generic formula (16) captures the sensitivity of mea-
sured inequality to different parts of the distribution, for any reference point
$e$. In the case of high values of $\alpha$ the index is particularly sensitive to high-
status inequality. For low and negative values of the parameter the opposite
is true. This is well known for the context in which one is using cardinal data
(income or wealth for example) in which case the mean is the appropriate
reference point and it is easy to show that $I_\alpha (s; \mu(s)) \geq 0$ for all $s$ and $\alpha$.
However in the context of ordinal data and where the reference point is not
the mean, more needs to be said.

Special Cases

First, it is necessary to be clear about the behaviour of the index $I_\alpha (s; e)$ in
two special cases of the sensitivity parameter,

**Theorem 4** For the class of measures $I_\alpha (s; e)$ given in (12) the limiting
cases, $\alpha = 0$ and $\alpha = 1$, are respectively equal to:

$$I_0 (s; e) = -\frac{1}{n} \sum_{i=1}^{n} \log s_i + \log e,$$

(17)

and

$$I_1 (s; e) = \begin{cases} \frac{1}{n} \sum_{i=1}^{n} s_i \log s_i - e \log e, & \text{if } e = \mu, \\ \pm \infty, & \text{if } e \neq \mu, \end{cases}$$

(18)

where $\mu = \frac{1}{n} \sum_{i=1}^{n} s_i$.

But (18) is not relevant for the analysis of ordinal data, for which the mean
is inappropriate as a reference point. In fact the behaviour of inequality for
$\alpha \geq 1$ is problematic, as we shall now show.
Negative values

For any $e \neq \mu(s)$ the index $I_\alpha$ can be negative for some values of $\alpha$ and is undefined for $\alpha = 1$. This behaviour is illustrated in Figure 2 for $e = 1$; this figure plots values of $I_\alpha(s;1)$ for different values of $\alpha$, using the distribution labeled Case 0 in Table 2. The problem with the index in the neighbourhood of $\alpha = 1$ is clear: $I_\alpha(s;1) \to +\infty$ when $\alpha \uparrow 1$ and $I_\alpha(s;1) \to -\infty$ when $\alpha \downarrow 1$.

Figure 2: The behaviour of $I_\alpha(s;1)$ as $\alpha$ varies
In the case where \( e = 1 \), using (17) we may rewrite (16) as

\[
I_\alpha(s, 1) = \begin{cases} 
\frac{1}{\alpha(\alpha-1)} \left[ \frac{1}{n} \sum_{i=1}^{n} s_i^{\alpha} - 1 \right], & \text{if } \alpha \neq 0,1, \\
-\frac{1}{n} \sum_{i=1}^{n} \log s_i. & \text{if } \alpha = 0.
\end{cases}
\]

(19)

### Parameter restriction

Here we focus principally on the ordinal data case with peer-inclusive status, for which it is appropriate to take the exogenous reference point \( e = 1 \). Clearly it would be appropriate to restrict the class of inequality measures so as to avoid the problems of negative values. To do this notice that \( I_\alpha(s, 1) \) can also be written

\[
I_\alpha(s, 1) = \frac{1}{\alpha - 1} \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{s_i^{\alpha} - 1}{\alpha} \right], \text{ if } \alpha \neq 0,1.
\]

(20)

For any \( \alpha \in \mathbb{R} \), it is clear that if \( 0 < s < 1 \) then \( [s^{\alpha} - 1]/\alpha < 0 \) and if \( s = 1 \) then \( [s^{\alpha} - 1]/\alpha = 0 \). So it is evident from (20) that \( I_\alpha(s; 1) \) is only well behaved under the parameter restriction \( \alpha < 1 \). But this parameter restriction is exactly the same as that required in order to obtain the Atkinson family of inequality indices from the family of generalised entropy indices. So our admissible class of inequality indices could also be written in the ordinally equivalent form

\[
A_\alpha(s) := \begin{cases} 
1 - \left[ \frac{1}{n} \sum_{i=1}^{n} s_i^{\alpha} \right]^{1/\alpha}, & \text{if } \alpha < 0 \text{ or } 0 < \alpha < 1, \\
1 - \left[ \prod_{i=1}^{n} s_i \right]^{1/n}, & \text{if } \alpha = 0.
\end{cases}
\]

(21)

### 4.4 Downward- or upward-looking status

Finally, let us examine the inequality outcomes if we use an upward-looking rather than downward-looking concept of status. The bottom two rows of Table 4 report the values of \( I_0 \) for the two concepts where the reference status is 1; in the upper part of this table the numbers of persons in each category are reproduced for reference (taken from Table 2). It comes as no surprise that the inequality outcome for each version of status is high in Case 1 and low in Case 2. For downward-looking status inequality is higher when the
<table>
<thead>
<tr>
<th>Case 0</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_k$</td>
<td>$n_k$</td>
<td>$n_k$</td>
<td>$n_k$</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>50</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>G</td>
<td>25</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>25</td>
<td>0</td>
</tr>
</tbody>
</table>

$I_0(s; 1)$  | 0.5198 | 0.5917 | 0.3465 | 0.4184 |
$I_0(s'; 1)$ | 0.4184 | 0.5917 | 0.3465 | 0.5198 |

Table 4: Inequality outcomes for downward-looking status ($s$) and upward-looking status ($s'$)

distribution is skewed towards the higher categories (Case 0); for upward-looking status inequality is higher when the distribution is skewed towards the lower categories (Case 3).

The reason for this is suggested by equation (19) where the number of active categories remains unchanged: the index decreases when the vector of status gets closer to a vector of 1. Label individuals in increasing order of utility. Then, from (2), we have $0 < s_1 \leq s_2 \leq \cdots \leq s_n \leq 1$ for downward-looking status: the status vector becomes closer to the vector unity when people move to the lower categories. For upward-looking status we have $1 \geq s'_1 \geq s'_2 \geq \cdots \geq s'_n > 0$: the status vector becomes closer to the vector unity when people move to the higher categories.

What happens if the number of active categories changes? For downward-looking status measures inequality must increase if a person migrates upwards from a category with multiple occupants to an empty category; for upward-looking status measures inequality must increase if a person migrates downwards from a category with multiple occupants to an empty category.\(^{20}\)

\(^{20}\)We can see this if we consider what happens to inequality when there is a merger. Suppose status is downward-looking, given by (2). Now let the non-empty categories $k^*$ and $k^*+1$ be merged. This is equivalent to increasing the size of category $k^*$ and emptying category $k^*+1$. For every $i$ such that $k(i) = k^*$ we see that $s_i$ increases by $n_{k^*+1}/n$; for every other $i$ status $s_i$ remains unchanged. So the status of some persons moves closer to $e$ and remains unchanged for others. By Axiom 2 this means that inequality must fall. By contrast, if a person migrates from category $k^*$ (where $n_{k^*} > 1$) to empty category $k^*+1$ this is the exact opposite of the merger just described: so inequality must rise. A similar argument can be constructed for upward-looking status.
5 Inequality measurement: practice

To be used in practice, we need to study the statistical properties of the inequality measure with ordinal data. In this section, we establish asymptotic distribution and finite sample performance of such inequality measures. Two empirical applications, in health and in happiness, illustrate the usefulness of the inequality measures in practice. In the light of the arguments in Section 4 we focus on ordered categorical variables, for which the appropriate reference point is the maximum status $e = 1$ and take the family of inequality indices (19). We consider samples with independent observations on $K$ ordered categories. Without loss of generality, any sample can be represented as follows:

$$x_i = \begin{cases} 
1 & \text{with sample proportion } p_1 \\
2 & \text{with sample proportion } p_2 \\
\vdots \\
K & \text{with sample proportion } p_K
\end{cases} \quad (22)$$

where $p_l$ is the number of observations in the $l$th category, divided by the sample size, so that $\sum_{l=1}^{K} p_l = 1$. For $K > 2$, the sample proportions $(p_1, p_2, \ldots, p_K)$ are known to follow a multinomial distribution with $n$ observations and a vector of probabilities $(\pi_1, \pi_2, \ldots, \pi_K)$. The status of observation $i$ is its position in the distribution, computed as the proportion of observations in the sample with a value less than or equal to $x_i$:

$$s_i = \frac{1}{n} \sum_{j=1}^{n} \iota(x_j \leq x_i) = \sum_{j=1}^{x_i} p_j \quad (23)$$

where $\iota(.)$ is the indicator function, equals to 1 if its argument is true and 0 otherwise. We provide the derivations for downward-looking status only, with status defined in (23). The extension to upward-looking status is straightforward: all one needs to do is to reverse the order of the categories.

5.1 Statistical properties

With a sample of categorical data, as defined in (22), and status given by the individual position, as defined in (23), we can rewrite the inequality measure
(19) as follows:

\[
I_\alpha = \begin{cases} 
\frac{1}{\alpha(\alpha-1)} \left[ \sum_{i=1}^{K} p_i \left[ \sum_{j=1}^{i} p_j \right]^{\alpha} - 1 \right] & \text{if } \alpha \neq 0,1, \\
- \sum_{i=1}^{K} p_i \log \left[ \sum_{j=1}^{i} p_j \right] & \text{if } \alpha = 0.
\end{cases}
\]

(24)

This measure is expressed as a non-linear function of \( K \) parameter estimates \((p_1, p_2, \ldots, p_K)\) following a multinomial distribution. From the Central Limit Theorem, \( I_\alpha \) follows asymptotically a Normal distribution with a covariance matrix which can be calculated by the delta method. Specifically, an estimator of the covariance matrix of \((p_1, p_2, \ldots, p_k)\) is given by

\[
\Sigma = \frac{1}{n} \begin{bmatrix}
    p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_K \\
    -p_2p_1 & p_2(1-p_2) & \cdots & -p_2p_K \\
    \vdots & \vdots & \ddots & \vdots \\
    -p_Kp_1 & -p_Kp_2 & \cdots & p_K(1-p_K)
\end{bmatrix}.
\]

(25)

The variance estimator for \( I_\alpha \) is equal to:

\[
\hat{\text{Var}}(I_\alpha) = D \Sigma D^T \quad \text{with} \quad D = \left[ \frac{\partial I_\alpha}{\partial p_1}; \frac{\partial I_\alpha}{\partial p_2}; \ldots; \frac{\partial I_\alpha}{\partial p_K} \right],
\]

(26)

where the \( l \)-th element of \( D \) is defined as

\[
\frac{\partial I_\alpha}{\partial p_l} = \begin{cases} 
\frac{1}{\alpha(\alpha-1)} \left( \left[ \sum_{i=1}^{l} p_i \right]^{\alpha} + \alpha \sum_{i=l+1}^{K-1} p_i \left[ \sum_{j=1}^{i} p_j \right]^{\alpha-1} \right) & \text{if } \alpha \neq 0,1, \\
- \log \left[ \sum_{j=1}^{l} p_j \right] - \sum_{i=l+1}^{K-1} p_i \left[ \sum_{j=1}^{i} p_j \right]^{-1} & \text{if } \alpha = 0.
\end{cases}
\]

(27)

For instance, in the case of three ordered categories, the inequality measures defined in (19) are equal to

\[
I_\alpha = \begin{cases} 
\frac{p_1^{\alpha+1} + p_2(p_1+p_2)^{\alpha} + p_3-1}{\alpha(\alpha-1)} & \text{if } \alpha \neq 0,1, \\
-p_1 \log p_1 - p_2 \log(p_1 + p_2) & \text{if } \alpha = 0,
\end{cases}
\]

where \( p_1, p_2 \) and \( p_3 \) are the proportions of observations in the respective ordered categories. Their variance estimators are given by (26), with \( \Sigma \)
defined in (25) where we use $K = 3$, and with $D$ equal to

$$D = \begin{cases} \frac{1}{\alpha(\alpha-1)} \left[ (\alpha+1)p_1^\alpha + \alpha p_2(p_1 + p_2)^{\alpha-1} ; (p_1 + p_2)^\alpha + \alpha p_2(p_1 + p_2)^{\alpha-1} ; 1 \right] & \text{if } \alpha \neq 0,1, \\ \left[ -\log p_1 - 1 - p_2/(p_1 + p_2) ; -\log(p_1 + p_2) - p_2/(p_1 + p_2) ; 0 \right] & \text{if } \alpha = 0. \end{cases}$$

We can use the variance estimators of $I_\alpha$ to compute test statistics and confidence intervals.

We now turn to the finite sample properties of the index. The coverage error rate of a confidence interval is the probability that the random interval does not include, or cover, the true value of the parameter. A method of constructing confidence intervals with good finite sample properties should provide a coverage error rate close to the nominal rate. For a confidence interval at 95%, the nominal coverage error rate is equal to 5%. We use Monte-Carlo simulation to approximate the coverage error rate of asymptotic confidence intervals in several experimental designs.

In our experiments, samples are drawn from a multinomial distribution with probabilities $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$. The status of observation $i$ is its position in the distribution, computed as the proportion of observations in the sample with a value less than or equal to $x_i$. For fixed values of $\alpha$, $n$, $K$ and $\pi$, we draw 10,000 samples. For each sample we compute $I_\alpha(s,1)$ and its confidence interval at 95%. The coverage error rate is computed as the proportion of times the true value of the inequality measure is not included in the confidence intervals.\footnote{The true values are $I_\alpha^{(0)} = \frac{1}{\alpha(\alpha-1)} \left[ \sum_{i=1}^{K} \pi_i \left[ \sum_{j=1}^{i} \pi_j \right]^\alpha - 1 \right] \alpha \neq 0$ and $I_0^{(0)} = -\sum_{i=1}^{K} \log \left[ \sum_{j=1}^{i} \pi_j \right]$.} Confidence intervals perform well in finite sample if the coverage error rate is close to the nominal value.

Table 5 shows coverage error rates of confidence intervals at 95% of $I_\alpha$ for different values of $\alpha = -1, 0, 0.5, 0.99, 1.01, 1.5, 2$, as the sample size increases, $n = 20, 50, 100, 200, 500, 1000$. We consider 3 ordered categories: samples are drawn from a multinomial distribution with probabilities $\pi = (0.3, 0.5, 0.2)$. If the asymptotic distribution is a good approximation of the exact distribution of the statistic, the coverage error rate should be close to the nominal error rate, 0.05. The results show that asymptotic confidence intervals perform well in finite sample, they are still reliable for $\alpha = 0.99, 1.01$ when the index is undefined for $\alpha = 1$.\footnote{The true values are $I_\alpha^{(0)} = \frac{1}{\alpha(\alpha-1)} \left[ \sum_{i=1}^{K} \pi_i \left[ \sum_{j=1}^{i} \pi_j \right]^\alpha - 1 \right] \alpha \neq 0$ and $I_0^{(0)} = -\sum_{i=1}^{K} \log \left[ \sum_{j=1}^{i} \pi_j \right]$.}
\[ \alpha \quad -1 \quad 0 \quad 0.5 \quad 0.99 \quad 1.01 \quad 1.5 \quad 2 \]

| \( n = 20 \) | 0.0606 | 0.0417 | 0.0598 | 0.0491 | 0.0491 | 0.0472 | 0.0431 |
| \( n = 50 \) | 0.0553 | 0.0518 | 0.0704 | 0.0601 | 0.0603 | 0.0483 | 0.0394 |
| \( n = 100 \) | 0.0499 | 0.0513 | 0.0684 | 0.0619 | 0.0619 | 0.0489 | 0.0416 |
| \( n = 200 \) | 0.0544 | 0.0476 | 0.0617 | 0.0613 | 0.0607 | 0.0543 | 0.0451 |
| \( n = 500 \) | 0.0523 | 0.0492 | 0.0521 | 0.0523 | 0.0526 | 0.0498 | 0.0466 |
| \( n = 1000 \) | 0.0485 | 0.0540 | 0.0552 | 0.0549 | 0.0551 | 0.0546 | 0.0528 |

Table 5: Coverage error rate of asymptotic confidence intervals at 95% of \( I_\alpha \), 10,000 replications, \( K = 3 \) and \( x \sim \text{Multinomial}(0.3, 0.5, 0.2) \).

5.2 Application

In the application, we use the data from the 5th wave of the World Values Survey 1981-2008, conducted in 2005-2008 over 56 countries.\(^{22}\) We focus our empirical study on two questions:

**Life satisfaction question:**

*All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means you are “completely dissatisfied” and 10 means you are “completely satisfied” where would you put your satisfaction with your life as a whole? (code one number):*

- Completely dissatisfied — 1 2 3 4 5 6 7 8 9 10 — Completely satisfied

**Health question:**

*All in all, how would you describe your state of health these days? Would you say it is (read out): 1 Very good, 2 Good, 3 Fair, 4 Poor.*

The relationship between life satisfaction and income has been studied extensively in the literature and is the subject of some disagreement.\(^{23}\) In a seminal paper, Easterlin (1974) showed that, for a given country, people with higher incomes are likely to report higher life satisfaction whereas, for cross-country comparisons and for higher-income countries, the average level of


\(^{23}\) See Clark and Senik (2011) for a recent survey.
life satisfaction does not vary much with higher income: this has come to be known as the Easterlin or happiness-income paradox. Several studies confirm the lack of impact of income on life satisfaction for higher-income countries, while other studies find a significant positive impact. To illustrate the issue, Figure 3 presents a cross-country comparison of the average of answers to the life satisfaction question (y-axis) and GDP per capita (x-axis). It is clear that the mean of life satisfaction is higher in countries with higher GDP per capita. However, the relationship is not linear and it is not obvious how to choose between a lognormal function and a piecewise linear regression model with the slope of the second line not significantly different from zero beyond $15,000 per head. So the presence or absence of income effect on life satisfaction among the higher income countries is not clear, and the controversy remains unresolved.

Many empirical studies make comparisons of the average of answers to questions, or they use specific transformations to calculate a composite index of well-being. In fact, they interpret the answers as cardinal with a linear scale, i.e., the values given to each successive answers are supposed to be equidistant. Such an interpretation is a matter of disagreement in the literature on Likert-type scale questionnaires. Some people consider that ordinal data provide information on ranks only and nothing else, so they should be treated as purely ordinal data with nonparametric statistics. For example, a horse-race result provides information on the rank order 1st, 2nd, 3rd, etc., without any information on the arrival time and differences in time between horses. Others consider that ordinal data can be interpreted as

---

26 GDP per capita in 2005 is measured in purchasing parity power chained dollars at 2005 constant prices, and comes from the Penn World Tables 7.0.
27 Layard (2003, p. 23) noted that “once a country has over $15,000 per head, its level of happiness appears to be independent of its income per head”. Deaton (2008) argues that different results are obtained from different datasets.
30 Another example is the IQ score, which can be used for rank comparisons, but not as interval data for many authors: the difference between an IQ of 130 and one of 100 is not equivalent to the difference between an IQ of 100 and one of 70 (Mackintosh 1998, Bartholomew 2004).
cardinal, with a linear scale; in some contexts – where the scale presents a
suitable symmetry of items around a clear middle category – the intervals
between points can be taken as approximately equal, so they can be analyzed
parametrically.\footnote{See Knapp (1990), Norman (2010).}

In our application, the main issue is whether we can assume that the
values assigned in each question can be treated as equidistant. In the health
question – Very Good, Good, Fair, Poor – the items are unlikely to be equidistant:
they are not symmetric (only one item can receive a below-average rating), and a bias would be introduced in favour of better outcomes. In the life
satisfaction question, the equidistance assumption implies that the distance
between ‘2’ and ‘3’ is similar to the distance between ‘5’ and ‘6’ or any other
two successive items. This position is quite common.\footnote{See Ng (1997), Ferrer-i-Carbonell and Frijters (2004) and Kristoffersen (2011) for a
discussion of this point.}

For our particular purpose, the relationship between the average of life satisfaction and GDP per capita in a cross-country comparison, we can illustrate how sensitive the results are to this hypothesis. Indeed, if we assume that life satisfaction is
highly correlated with personal incomes and that the distribution of incomes
is lognormal, it makes sense to represent the answers on an exponential scale
rather than on a linear scale.\footnote{Here, we assume that everybody use the same scale, which is also questionable.}

Figure 4 presents a cross-country comparison of the average of the exponential transformation of the code numbers from
the life satisfaction question (y-axis) and GDP per capita (x-axis). There
is no clear relationship between the average of life satisfaction and national
income. Column (i) in Table 6 shows OLS regression results of the average
of life satisfaction (on the exponential scale) on GDP per capita. The slope
coefficient is not significantly different from zero and so we may conclude
that there is no significant relationship between life satisfaction and GDP
per capita. Compared to the conclusions drawn from Figure 3 and based
on a linear scale, we obtain very different results. Clearly, the results are
sensitive to the cardinal interpretation of the answers.

A simple solution might seem to be to replace the mean by the median
in the empirical analysis. However, the median is not well-defined for ordinal
data, in particular when there is a small number of categories (see the
discussion in section 4.2 and footnote 19). Another solution might be to use
the fraction of people with a score higher than \( c \). However, in our example,
different choices of \( c \) produce different results: the relationship between the
<table>
<thead>
<tr>
<th></th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>4079</td>
<td>0.7358</td>
<td>0.5245</td>
<td>0.3450</td>
</tr>
<tr>
<td></td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
</tr>
<tr>
<td>GDP per capita</td>
<td>0.0225</td>
<td>-1.53e-06</td>
<td>-1.03e-06</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.292)</td>
<td>(0.008)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>$I_0$(LifeSatis)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.2274</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.019)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02</td>
<td>0.12</td>
<td>0.11</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 6: Cross-country regressions, with $p$-values in parenthesis. Column (i) shows the OLS estimation of the average of life satisfaction (on the exponential scale) on GDP per capita; column (ii) the estimation of the inequality of life satisfaction, $I_0$(LifeSatis), on GDP per capita; column (iii) the estimation of the inequality of health, $I_0$(health), on GDP per capita; and column (iv) the estimation of $I_0$(health) on $I_0$(LifeSatis).

The fraction of people with an answer higher than $c = 5$ and GDP per capita is significantly positive from an OLS regression, while it is not significant with $c = 8$ (results not reported). Hereafter, we use the index developed in this paper, which avoids all such problems.

The inequality index developed in this paper is defined on individual ranks alone and is thus insensitive to any cardinal interpretation of the answers. Whether an upward-looking or downward-looking status concept is appropriate will depend on the context. If we require inequality to fall if answers are skewed toward more desirable categories, then it is clear that, for the life satisfaction question, where the categories are (Completely dissatisfied - 1, 2, ..., 10 - Completely satisfied) an upward-looking version is required. For the health question the lowest category is assigned to the “very good” answer; a downward-looking version ensures that inequality decreases when people tend to report better health states.

If GDP were to be strongly related to life satisfaction, then one would not expect any specific relationship between GDP and the dispersion of life...
satisfaction.\footnote{To illustrate, consider the model \( \text{LifeSatisf}_i = \alpha + \beta \text{GDP}_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2) \), where we assume that \( \beta \) is a significant coefficient and the \( R^2 \) is large: there is a strong (linear) relationship between LifeSatisf and GDP. If in addition the LifeSatisf equation is homoskedastic (\( \sigma_i^2 \) constant), then this suggests that there is no relationship between GDP and the dispersion of LifeSatisf; whatever the value of GDP, the dispersion of LifeSatisf is the same. If the LifeSatisf equation is heteroskedastic (\( \sigma_i^2 \) is non-constant), the dispersion of LifeSatisf may or may not be related to GDP: in a regression model, the form of the heteroskedasticity cannot be deduced from the relationship between the dependent variable and the covariate. However, if every country has a different value of GDP, \( \sigma_i^2 \) measures the dispersion of LifeSatisf for a given country taking the measure \( I_0 \) as a measure of dispersion, the same reasoning applies with \( I(\text{LifeSatisf}) \) replacing \( \sigma_i^2 \).} Figures 5 and 6 present a cross-country comparison of the inequality index of life satisfaction\footnote{The inequality index, \( I_0 \), is computed with \( \alpha = 0 \), as defined in (24).} with GDP per capita (Figure 5) and income inequality (Figure 6). In Figure 5 a negative relationship seems to emerge, but it is far from clear-cut: an OLS regression of the inequality index of life satisfaction, \( I_0(\text{LifeSatisf}) \), on GDP per capita shows that the slope coefficient is significantly different from zero at 1\% (see Table 6, column (ii)). However, as the \( p \)-value is very close to 0.01 and the \( R^2 = 0.12 \) is small, it suggests that GDP per capita is not strongly related to the inequality of life satisfaction. In other words, people in high-income countries are not necessarily those who tend to report higher life satisfaction. Indeed, we can see from Figure 5 that the two countries with the lowest inequality index, Colombia and Mexico, have very small values of GDP per capita. When we look at the detailed data, presented in Table 7, we can see that more than 73.9\% of the individuals in Colombia and more than 77.1\% of the individuals in Mexico, report values higher or equal to ‘8’ in the original life satisfaction question. The figure suggests that, by including or excluding certain countries, we could have more or less significant results on the happiness-income relationship. Nevertheless, it is clear from those results that the happiness-income relationship is weak in cross-country comparisons.\footnote{We obtain similar results with different values of \( \alpha \), the \( R^2 \) is slightly higher with \( \alpha = 0.5 \) and slightly lower with \( \alpha = -0.5, -1 \).}

We now turn to the responses to the health question, detailed in Table 8.\footnote{Over all the countries, Guatemala, Turkey, Uruguay and South Africa report a few values equal to 5, associated with a “Very poor” answer. From the \textit{Documentation of the Values Surveys}, for the other countries, the “Very poor” is not a possible answer in the questionnaire. We thus remove those four countries from the database.} When we look at the relationship between the fraction of people \textit{satisfied} with their health – answering “Very good” or “Good” – and GDP per capita, we find
a significant positive relationship, as obtained in Deaton (2008). But, when we look at the fraction of people very satisfied with their health — answering "Very good" only — the relationship is not significant from an OLS estimation. Once again, the results appear to be sensitive to the fraction of people used (with a code number less than ‘c’=2 or ‘c’=1). It may be explained by some countries, which behave very differently when we consider satisfied and very satisfied people. For instance, Hong Kong and Rwanda have, respectively, 63.5% and 33.7% of satisfied people, but only 5.6% and 2.5% of very satisfied people (see Table 8). The index developed in this paper takes into account the cumulative mass of individuals to each answer; for downward-looking status is defined to decrease when the distribution of answers gets closer to the case where everybody give the answer associated with the lowest value ("Very good" for the health question). In the following, we use this index to study health-income and health-life satisfaction relationships.

Figures 7 and 8 present a cross-country comparison of the inequality index of health with GDP per capita (Figure 7) and income inequality (Figure 8). There is no clear relationship, and the OLS estimation of $I_0(\text{health})$ on GDP per capita produces a slope coefficient not significantly different from zero at 1%, and a $R^2 = 0.11$ (see Table 6, column (iii)). These results suggest that the health-income relationship is not significant. In other words, people in higher income countries do not tend to report higher health satisfaction.

Figure 9 presents a cross-country comparison of the inequality index of health and the inequality of life satisfaction. Once again, there is no clear relationship, and the OLS estimation of $I_0(\text{health})$ on $I_0(\text{LifeSatisf})$ produces a slope coefficient not significantly different from zero at 1%, and a $R^2 = 0.10$ (see Table 6, column (iv)). These results suggest that the health-life satisfaction relationship is not significant, that is, countries where people tend to report higher life satisfaction are not necessarily those where people tend to report higher health satisfaction.

6 Conclusion

We can provide a precise answer to the problem of measuring inequality of the distribution of ordinal data interpreted as categorical variables.

---

38 Again the inequality index, $I_\alpha$, is computed with $\alpha = 0$, as defined in (24).
39 We obtain similar results with $\alpha = -1, -0.5, 0.5$.  

31
There are three basic ingredients: the concept of status within a distribution, a reference point and a set of axioms. Status can be downward- or upward-looking depending on the context of the analysis. The axiomatisation is general and provides an approach for a variety of data types including the case of categorical data – it characterises a family of indices that is conditional on a sensitivity parameter and a reference point. The specific class of inequality measures that emerges from the axiomatisation is related to the Generalised Entropy and Atkinson classes; but, by contrast to conventional inequality analysis, the reference point for categorical data is not the mean of the distribution but the maximum possible value of status.

The approach is straightforward to implement empirically. As we have shown in section 5.2, how you treat categorical data in empirical studies really matters: if we were to treat ordinal variables as though they were cardinal the cardinalisation that is imposed affects conclusions about whether there is a relationship between inequality and income.
References


<table>
<thead>
<tr>
<th>Country</th>
<th>Life satisfaction question</th>
<th>Inequality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>14 19 15 15 15 15 15 15</td>
<td>0.638 0.00077</td>
</tr>
<tr>
<td>Australia</td>
<td>17 18 20 40 109 117 303 472 195 110</td>
<td>0.689 0.00065</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>55 71 103 186 399 233 171 126 48 107</td>
<td>0.755 0.00049</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>57 69 115 116 184 109 147 81 49 30</td>
<td>0.804 0.00032</td>
</tr>
<tr>
<td>Brazil</td>
<td>29 15 21 34 162 130 185 352 201 366</td>
<td>0.653 0.00094</td>
</tr>
<tr>
<td>Canada</td>
<td>12 17 25 52 123 158 361 726 348 335</td>
<td>0.663 0.00064</td>
</tr>
<tr>
<td>Switzerland</td>
<td>8 7 12 16 64 78 170 445 241 191</td>
<td>0.647 0.00059</td>
</tr>
<tr>
<td>Chile</td>
<td>11 3 24 56 120 125 197 317 130 165</td>
<td>0.710 0.00060</td>
</tr>
<tr>
<td>China</td>
<td>72 72 86 85 222 247 242 427 198 275</td>
<td>0.731 0.00068</td>
</tr>
<tr>
<td>Colombia</td>
<td>44 20 20 40 172 157 332 601 147 203</td>
<td>0.553 0.00123</td>
</tr>
<tr>
<td>Cyprus</td>
<td>28 8 17 25 84 166 324 335 170 69</td>
<td>0.696 0.00052</td>
</tr>
<tr>
<td>Egypt</td>
<td>305 122 234 251 510 311 170 445 241 191</td>
<td>0.689 0.00065</td>
</tr>
<tr>
<td>France</td>
<td>8 7 12 16 64 78 170 445 241 191</td>
<td>0.681 0.00039</td>
</tr>
<tr>
<td>France</td>
<td>30 22 47 88 172 219 246 83 45 74</td>
<td>0.748 0.00057</td>
</tr>
<tr>
<td>Morocco</td>
<td>27 31 115 195 373 206 130 58 22 40</td>
<td>0.761 0.00057</td>
</tr>
<tr>
<td>Moldova</td>
<td>46 69 99 150 172 151 131 127 66 30</td>
<td>0.800 0.00030</td>
</tr>
<tr>
<td>Mexico</td>
<td>39 10 15 18 86 60 117 380 256 531</td>
<td>0.565 0.00113</td>
</tr>
<tr>
<td>Mali</td>
<td>113 31 79 95 328 154 184 148 96 202</td>
<td>0.749 0.00072</td>
</tr>
<tr>
<td>Malaysia</td>
<td>17 9 26 41 145 230 287 365 92 86</td>
<td>0.729 0.00049</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2 5 5 16 40 82 235 406 152 102</td>
<td>0.715 0.00061</td>
</tr>
<tr>
<td>Norway</td>
<td>30 22 47 88 172 219 246 83 45 74</td>
<td>0.748 0.00057</td>
</tr>
<tr>
<td>New Zealand</td>
<td>7 8 16 20 50 67 149 224 182 204</td>
<td>0.638 0.00058</td>
</tr>
<tr>
<td>Peru</td>
<td>29 24 59 72 194 165 278 280 111 278</td>
<td>0.703 0.00082</td>
</tr>
<tr>
<td>Poland</td>
<td>14 15 34 34 152 117 149 238 112 124</td>
<td>0.723 0.00038</td>
</tr>
<tr>
<td>Romania</td>
<td>93 82 163 143 283 179 247 285 107 76</td>
<td>0.785 0.00045</td>
</tr>
<tr>
<td>Russia</td>
<td>117 74 104 151 362 217 326 342 146 177</td>
<td>0.768 0.00049</td>
</tr>
<tr>
<td>Rwanda</td>
<td>91 146 136 166 365 299 122 96 53 29</td>
<td>0.790 0.00049</td>
</tr>
<tr>
<td>Serbia</td>
<td>30 53 84 97 177 170 262 210 54 38</td>
<td>0.760 0.00062</td>
</tr>
<tr>
<td>Slovakia</td>
<td>10 6 19 25 167 122 173 240 104 167</td>
<td>0.702 0.00068</td>
</tr>
<tr>
<td>Sweden</td>
<td>5 5 16 22 47 61 188 367 177 114</td>
<td>0.661 0.00060</td>
</tr>
<tr>
<td>Thailand</td>
<td>12 11 36 53 137 217 307 429 172 158</td>
<td>0.710 0.00056</td>
</tr>
<tr>
<td>Trinidad&amp;Tobago</td>
<td>35 7 14 29 120 124 176 195 80 219</td>
<td>0.679 0.00091</td>
</tr>
<tr>
<td>Turkey</td>
<td>41 18 46 36 77 134 220 291 211 272</td>
<td>0.681 0.00080</td>
</tr>
<tr>
<td>Taiwan</td>
<td>38 16 34 54 190 207 217 280 74 117</td>
<td>0.734 0.00059</td>
</tr>
<tr>
<td>Ukraine</td>
<td>45 51 97 81 147 147 170 149 66 43</td>
<td>0.787 0.00038</td>
</tr>
<tr>
<td>Uruguay</td>
<td>19 2 12 19 76 125 218 245 115 162</td>
<td>0.689 0.00065</td>
</tr>
<tr>
<td>United States</td>
<td>9 11 31 54 91 117 275 358 224 71</td>
<td>0.698 0.00061</td>
</tr>
<tr>
<td>Vietnam</td>
<td>6 10 41 61 163 287 278 292 138 206</td>
<td>0.719 0.00039</td>
</tr>
<tr>
<td>South Africa</td>
<td>96 75 97 141 257 298 419 588 435 571</td>
<td>0.699 0.00081</td>
</tr>
<tr>
<td>Zambia</td>
<td>107 52 89 114 204 189 249 229 99 131</td>
<td>0.773 0.00030</td>
</tr>
</tbody>
</table>

Table 7: Number of people per code number in the life satisfaction question (source: World Values Surveys, 5th wave) and the inequality index with its standard error (upward looking status).
## Table 8: Number of people per code number in the health question (*source*: World Values Surveys, 5th wave) and the inequality index (downward-looking status) with its standard error.

<table>
<thead>
<tr>
<th>Health question</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>325</td>
<td>465</td>
<td>186</td>
<td>27</td>
<td>0.480</td>
<td>0.00069</td>
</tr>
<tr>
<td>Australia</td>
<td>402</td>
<td>675</td>
<td>274</td>
<td>61</td>
<td>0.496</td>
<td>0.00068</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>427</td>
<td>670</td>
<td>350</td>
<td>72</td>
<td>0.511</td>
<td>0.00065</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>149</td>
<td>389</td>
<td>310</td>
<td>150</td>
<td>0.575</td>
<td>0.00062</td>
</tr>
<tr>
<td>Brazil</td>
<td>408</td>
<td>700</td>
<td>355</td>
<td>36</td>
<td>0.501</td>
<td>0.00061</td>
</tr>
<tr>
<td>Canada</td>
<td>857</td>
<td>880</td>
<td>341</td>
<td>81</td>
<td>0.461</td>
<td>0.00086</td>
</tr>
<tr>
<td>Switzerland</td>
<td>400</td>
<td>635</td>
<td>173</td>
<td>33</td>
<td>0.462</td>
<td>0.00068</td>
</tr>
<tr>
<td>Chile</td>
<td>106</td>
<td>475</td>
<td>277</td>
<td>52</td>
<td>0.524</td>
<td>0.00067</td>
</tr>
<tr>
<td>China</td>
<td>119</td>
<td>617</td>
<td>486</td>
<td>290</td>
<td>0.550</td>
<td>0.00080</td>
</tr>
<tr>
<td>Colombia</td>
<td>674</td>
<td>1517</td>
<td>670</td>
<td>72</td>
<td>0.502</td>
<td>0.00065</td>
</tr>
<tr>
<td>Cyprus</td>
<td>413</td>
<td>387</td>
<td>184</td>
<td>65</td>
<td>0.478</td>
<td>0.00092</td>
</tr>
<tr>
<td>Egypt</td>
<td>471</td>
<td>1578</td>
<td>766</td>
<td>236</td>
<td>0.515</td>
<td>0.00090</td>
</tr>
<tr>
<td>Spain</td>
<td>237</td>
<td>716</td>
<td>195</td>
<td>48</td>
<td>0.463</td>
<td>0.00091</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>386</td>
<td>582</td>
<td>390</td>
<td>139</td>
<td>0.544</td>
<td>0.00063</td>
</tr>
<tr>
<td>Finland</td>
<td>233</td>
<td>437</td>
<td>284</td>
<td>61</td>
<td>0.534</td>
<td>0.00059</td>
</tr>
<tr>
<td>France</td>
<td>300</td>
<td>418</td>
<td>225</td>
<td>58</td>
<td>0.513</td>
<td>0.00060</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>345</td>
<td>412</td>
<td>196</td>
<td>87</td>
<td>0.508</td>
<td>0.00080</td>
</tr>
<tr>
<td>Georgia</td>
<td>189</td>
<td>439</td>
<td>574</td>
<td>296</td>
<td>0.600</td>
<td>0.00042</td>
</tr>
<tr>
<td>Germany</td>
<td>436</td>
<td>908</td>
<td>538</td>
<td>148</td>
<td>0.534</td>
<td>0.00063</td>
</tr>
<tr>
<td>Ghana</td>
<td>571</td>
<td>661</td>
<td>208</td>
<td>93</td>
<td>0.471</td>
<td>0.00083</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>68</td>
<td>733</td>
<td>407</td>
<td>48</td>
<td>0.435</td>
<td>0.00139</td>
</tr>
<tr>
<td>Indonesia</td>
<td>386</td>
<td>1144</td>
<td>408</td>
<td>63</td>
<td>0.477</td>
<td>0.00086</td>
</tr>
<tr>
<td>India</td>
<td>438</td>
<td>934</td>
<td>489</td>
<td>136</td>
<td>0.526</td>
<td>0.00068</td>
</tr>
<tr>
<td>Iran</td>
<td>492</td>
<td>1294</td>
<td>680</td>
<td>156</td>
<td>0.520</td>
<td>0.00074</td>
</tr>
<tr>
<td>Iraq</td>
<td>476</td>
<td>1285</td>
<td>592</td>
<td>317</td>
<td>0.536</td>
<td>0.00081</td>
</tr>
<tr>
<td>Italy</td>
<td>188</td>
<td>562</td>
<td>228</td>
<td>34</td>
<td>0.487</td>
<td>0.00084</td>
</tr>
<tr>
<td>Jordan</td>
<td>524</td>
<td>518</td>
<td>130</td>
<td>27</td>
<td>0.425</td>
<td>0.00086</td>
</tr>
<tr>
<td>Japan</td>
<td>148</td>
<td>440</td>
<td>405</td>
<td>95</td>
<td>0.554</td>
<td>0.00063</td>
</tr>
<tr>
<td>Korea (Republic of)</td>
<td>173</td>
<td>809</td>
<td>208</td>
<td>8</td>
<td>0.416</td>
<td>0.00115</td>
</tr>
<tr>
<td>Morocco</td>
<td>461</td>
<td>434</td>
<td>262</td>
<td>43</td>
<td>0.482</td>
<td>0.00086</td>
</tr>
<tr>
<td>Moldova</td>
<td>107</td>
<td>443</td>
<td>354</td>
<td>117</td>
<td>0.547</td>
<td>0.00086</td>
</tr>
<tr>
<td>Mexico</td>
<td>344</td>
<td>654</td>
<td>503</td>
<td>54</td>
<td>0.532</td>
<td>0.00049</td>
</tr>
<tr>
<td>Mali</td>
<td>437</td>
<td>510</td>
<td>467</td>
<td>93</td>
<td>0.536</td>
<td>0.00064</td>
</tr>
<tr>
<td>Malaysia</td>
<td>332</td>
<td>739</td>
<td>125</td>
<td>5</td>
<td>0.426</td>
<td>0.00066</td>
</tr>
<tr>
<td>Netherlands</td>
<td>247</td>
<td>497</td>
<td>262</td>
<td>42</td>
<td>0.513</td>
<td>0.00063</td>
</tr>
<tr>
<td>New Zealand</td>
<td>435</td>
<td>388</td>
<td>165</td>
<td>46</td>
<td>0.460</td>
<td>0.00093</td>
</tr>
<tr>
<td>New Zealand</td>
<td>348</td>
<td>427</td>
<td>154</td>
<td>20</td>
<td>0.462</td>
<td>0.00074</td>
</tr>
<tr>
<td>Peru</td>
<td>112</td>
<td>606</td>
<td>723</td>
<td>59</td>
<td>0.511</td>
<td>0.00085</td>
</tr>
<tr>
<td>Poland</td>
<td>175</td>
<td>365</td>
<td>331</td>
<td>127</td>
<td>0.575</td>
<td>0.00051</td>
</tr>
<tr>
<td>Romania</td>
<td>175</td>
<td>776</td>
<td>569</td>
<td>254</td>
<td>0.551</td>
<td>0.00092</td>
</tr>
<tr>
<td>Russia</td>
<td>134</td>
<td>731</td>
<td>911</td>
<td>296</td>
<td>0.548</td>
<td>0.00086</td>
</tr>
<tr>
<td>Rwanda</td>
<td>37</td>
<td>469</td>
<td>740</td>
<td>255</td>
<td>0.523</td>
<td>0.00103</td>
</tr>
<tr>
<td>Serbia</td>
<td>235</td>
<td>431</td>
<td>406</td>
<td>147</td>
<td>0.574</td>
<td>0.00049</td>
</tr>
<tr>
<td>Slovenia</td>
<td>220</td>
<td>369</td>
<td>302</td>
<td>144</td>
<td>0.574</td>
<td>0.00054</td>
</tr>
<tr>
<td>Sweden</td>
<td>333</td>
<td>430</td>
<td>187</td>
<td>33</td>
<td>0.480</td>
<td>0.00075</td>
</tr>
<tr>
<td>Thailand</td>
<td>331</td>
<td>834</td>
<td>295</td>
<td>66</td>
<td>0.488</td>
<td>0.00079</td>
</tr>
<tr>
<td>Trinidad&amp;Tobago</td>
<td>321</td>
<td>375</td>
<td>248</td>
<td>56</td>
<td>0.515</td>
<td>0.00073</td>
</tr>
<tr>
<td>Taiwan</td>
<td>280</td>
<td>772</td>
<td>117</td>
<td>58</td>
<td>0.439</td>
<td>0.00086</td>
</tr>
<tr>
<td>Ukraine</td>
<td>76</td>
<td>412</td>
<td>378</td>
<td>132</td>
<td>0.545</td>
<td>0.00094</td>
</tr>
<tr>
<td>United States</td>
<td>350</td>
<td>615</td>
<td>206</td>
<td>47</td>
<td>0.480</td>
<td>0.00069</td>
</tr>
<tr>
<td>Vietnam</td>
<td>155</td>
<td>736</td>
<td>473</td>
<td>131</td>
<td>0.519</td>
<td>0.00101</td>
</tr>
<tr>
<td>Zambia</td>
<td>415</td>
<td>618</td>
<td>282</td>
<td>128</td>
<td>0.520</td>
<td>0.00072</td>
</tr>
</tbody>
</table>
Figure 3: Cross-country comparison of average life satisfaction and GDP per capita.
Figure 4: Cross-country comparison of average life satisfaction and GDP per capita, using an exponential scale for values assigned to the answers.
Figure 5: Cross-country comparison of inequality of life satisfaction and GDP per capita.
Figure 6: Cross-country comparison of life-satisfaction inequality and income inequality.
Figure 7: Cross-country comparison of health inequality and GDP per capita.
Figure 8: Cross-country comparison of health inequality and income inequality.
Figure 9: Cross-country comparison of inequality of health and inequality of life satisfaction
Appendix: Proofs

Proof. [Theorem 1] There are two cases to consider. (1) In the case where the data are categorical $S$ is the set of non-negative rational numbers, $\mathbb{Q}_+$. (2) In the case where the data have cardinal significance $S$ can be taken as an interval in $\mathbb{R}$. In either case $(S, +, \succ)$ forms a strictly ordered group (Krantz 1964, Luce and Tukey 1964, Wakker 1988) and so, from Theorem 5.3 of Fishburn (1970) Axioms 1-3 jointly imply that, for a given $e$, $\succeq$ is representable by a continuous function $S^{n+1} \to \mathbb{R}$:

$$\sum_{i=1}^{n} d_i(s_i, e), \forall (s, e) \in S^{n+1}$$  \hspace{1cm} (28)

where, for each $i$, $d_i : S \to \mathbb{R}$ is a continuous function. By Axiom 2 this is increasing in $s_i$ if $s_i > e$ and *vice versa*. In view of Axiom 4 the functions $d_i$ must all be identical. Clearly the ordering $\succeq$ is also representable any monotonic transform of the function in (28), possibly depending on $e$, and so the result follows. ■

Proof. [Theorem 2] From (7) $(s, e) \sim (s', e')$ implies

$$\sum_{i=1}^{n} d(s_i, e) = \sum_{i=1}^{n} d(s'_i, e')$$  \hspace{1cm} (29)

and so Axiom 5 implies

$$\sum_{i=1}^{n} d(\lambda s_i, e) = \sum_{i=1}^{n} d(\lambda s'_i, e').$$

Therefore we have

$$\frac{\sum_{i=1}^{n} d(\lambda s_i, e)}{\sum_{i=1}^{n} d(s_i, e)} = \frac{\sum_{i=1}^{n} d(\lambda s'_i, e')}{\sum_{i=1}^{n} d(s'_i, e')}.$$  \hspace{1cm} (30)

For any $\lambda > 0$ we have

$$\sum_{i=1}^{n} d(\lambda s_i, e) = \theta \left( \lambda, \sum_{i=1}^{n} d(s_i, e) \right),$$  \hspace{1cm} (31)

where $\theta : \mathbb{R} \to \mathbb{R}$ is increasing in its second argument. Consider the case where, for arbitrary distinct values $j$ and $k$, we have $s_i = e$ for all $i \neq j, k$.  

46
This implies that, for given values of \(e, \lambda\), (31) can be written as the functional equation:
\[
f (s_j) + f (s_k) = h (g (s_j) + g (s_k)), \tag{32}
\]
where
\[
f (s) := d (\lambda s, e) + \left[ \frac{1}{2} n - 1 \right] d (\lambda e, e), \tag{33}
g (s) := d (s, e) + \left[ \frac{1}{2} n - 1 \right] d (e, e) \tag{34}
\]
and \(h (x) := \theta (\lambda, x)\). Equation (32) has the solution
\[
f (s) = a_0 g (s) + a_1, \tag{35}
h (x) = a_0 x + 2a_1, \tag{36}
\]
where \(a_0, a_1\) are constants (Polyanin and Zaitsev 2004, Supplement S.5.5) that may depend on the arbitrary values of \(e\) and \(\lambda\). Using (35) and (33) in the case \(s = e\) we have:
\[
a_1 = \frac{1}{2} n [d (\lambda e, e) - a_0 d (e, e)]
\]
Using this in (35) and (33) and writing \(a_0 = \psi (\lambda, e)\) we have
\[
d (\lambda s, e) = \psi (\lambda, e) d (s, e) + \frac{1}{2} n [d (\lambda e, e) - \psi (\lambda, e) d (e, e)]
\]
and so
\[
d (\lambda e, e) = \psi (\lambda, e) d (e, e)
\]
from which we may deduce
\[
d (\lambda s, e) = \psi (\lambda, e) d (s, e). \tag{37}
\]
Substituting from (37) into (30) we find that \(\psi (\lambda, e) = \psi (\lambda, e')\): therefore \(\psi\) is independent of \(e\).
\[
d (\lambda s, e) = \psi (\lambda) d (s, e). \tag{38}
\]
For given \(e\) we may write (38) as
\[
\theta_1 (u + v) = \theta_1 (u) + \theta_2 (v)
\]
where \( u := \log s \), \( v := \log \lambda \), \( \theta_1 (u) := \log d \left( \exp (u), e \right) \), \( \theta_2 (v) := \log \psi \left( \exp (v) \right) \). The solution to this equation is \( \theta_1 (u) = \alpha u + \beta \), \( \theta_2 (v) = \alpha v \) and so equation (38) implies that
\[
d (s, e) = A s^\alpha,
\]
\[
\psi (\lambda) = \lambda^\alpha,
\]
where \( \alpha \) and \( A \) are constants that may depend on the exogenous value of \( e \).

**Outline Proof. [Theorem 3]** The proof follows the proof of Theorem 2 from (29) to (37) closely, substituting \((\lambda s, \lambda e)\) for \((s, e)\). The counterpart of equation (37) is
\[
d (\lambda s, \lambda e) = \psi (\lambda, e) d (s, e) .
\]
and once again we find that \( \psi \) is independent of \( e \). So
\[
d (\lambda s, \lambda e) = \psi (\lambda) d (s, e).
\]

From Aczél and Dhombres (1989), page 346, there must exist \( \beta \in \mathbb{R} \) and a function \( \phi : \mathbb{R}_+ \rightarrow \mathbb{R} \) such that (9) is satisfied.

**Proof. [Theorem 4]** L’Hôpital’s rule states that the limit of an undefined ratio between two functions of the same variable is equal to the limit of the ratio of their first derivative. Let us define \( f (\alpha) = \frac{1}{n} \sum_{i=1}^{n} s_i^\alpha - e^\alpha = \frac{1}{n} \sum_{i=1}^{n} \exp (\alpha \log s_i) - \exp (\alpha \log e) \) and \( g (\alpha) = \alpha (\alpha - 1) \), from which we have \( f' (\alpha) = \frac{1}{n} \sum_{i=1}^{n} s_i^\alpha \log s_i - e^\alpha \log e \) and \( g' (\alpha) = 2 \alpha - 1 \). For \( \alpha = 0 \), it follows from a simple application of l’Hôpital’s rule that,
\[
I_0 (s; e) = \lim_{\alpha \to 0} \frac{f (\alpha)}{g (\alpha)} = \frac{f' (0)}{g' (0)} = - \left( \frac{1}{n} \sum_{i=1}^{n} \log s_i - \log e \right)
\]

L’Hôpital’s rule applies when the ratio \( f (\alpha) / g (\alpha) \) is undefined only. For the case \( \alpha = 1 \), this ratio is undefined only if \( e = \frac{1}{n} \sum_{i=1}^{n} s_i = \mu \), otherwise the ratio is defined and is equal to \( \pm \infty \). Indeed, it is straightforward to see that \( f (1) = 0 \) if \( e = \mu \), otherwise \( f (1) \neq 0 \). It follows that, for \( \alpha = 1 \), l’Hôpital’s rule applies only when \( e = \mu \):
\[
I_1 (s; e = \mu) = \lim_{\alpha \to 1} \frac{f (\alpha)}{g (\alpha)} = \frac{f' (1)}{g' (1)} = \frac{1}{n} \sum_{i=1}^{n} s_i \log s_i - e \log e.
\]