Reference Distributions and Inequality Measurement

Frank A. Cowell,[1] Emmanuel Flachaire[2]
and Sanghamitra Bandyopadhyay[3]

May 2012

[1]Sticerd, London School of Economics, Houghton Street, London WC2A 2AE
[2]Greqam, Aix-Marseille Université, 2 rue de la Charité 13002 Marseille
[3]Queen Mary, University of London, Mile End Road, London E1 4NS

Abstract

We investigate the general problem of comparing pairs of distribution which includes approaches to inequality measurement, the evaluation of "unfair" income inequality, evaluation of inequality relative to norm incomes, and goodness of fit. We show how information theory can be used to provide insights on the problem and characterise a class of divergence measures using a parsimonious set of axioms. The problems of appropriate statistical implementation are discussed and empirical illustrations of the technique are provided using a variety of reference distributions.

- Keywords: generalised entropy measures, income distribution, inequality measurement

- JEL Classification: D63, C10

1 Introduction

There is a broad class of problems in distributional analysis that involve comparing two distributions. This may involve judging whether a functional form is a good fit to an empirical distribution; it may involve computation of the divergence of an empirical distribution from a theoretical economic model; it may involve the ethical evaluation of an empirical distribution with reference to some norm or ideal distribution. This paper shows how the class of problems can be characterised in a way that has a natural interpretation in terms of familiar analytical tools.

This is not some recondite or abstruse topic. Several authors have explicitly characterised inequality using this two-distribution paradigm: an inequality measure is defined in terms of the divergence of an empirical income distribution from an equitable reference distribution.[1] Furthermore several recent papers have revived interest in the idea of inequality evaluations with reference to "norm incomes" or a reference distribution;[2] in particular some authors have focused on replacing a perfectly egalitarian reference distribution with one that takes explicit account of fairness.[3] In recent contributions Magdalou and Nock (2011) have also examined the concept of divergence between any two income distributions and its economic interpretation and Cowell et al. (2011) show how similar concepts can be used to formulate an approach to the measurement of goodness of fit.[4]

Our approach to the problem is based on standard results in information theory that allow one to construct a distance concept that is appropriate for characterising the divergence between the empirical distribution function and a proposed reference distribution. The connection between information theory and the economic interpretation of distributions is established by exploiting the close relationship between entropy measures (based on probability distributions) and measures of inequality and distributional change (based on distributions of income shares). The approach is adaptable to other fields in economics that make use of models of distributions. The paper is structured as follows. In Section 2 we explain the connection be-

[1] See, for example, Bartels (1977) and Nygård and Sandström (1981). See also Ebert (1984)'s characterisation of absolute inequality indices in ternms of distance between income distributions.

[2] Almås et al. (2011) compare "actual and equalizing earnings;" their work is related to Paglin's Gini (Paglin 1975) and Wertz's Gini (Wertz 1979). Also see Jenkins and O'Higgins (1989) and Garvy (1952).

[3] On the fairness reference distribution see, for example Almås et al. (2011) and Devooght (2008).

[4] The Cowell et al. (2011) approach differs from that developed here in that it deals with the problem of continuous reference distributions on unbounded support.

tween information theory and the analysis of income distributions. Section 3 introduces different concepts of reference distribution that are relevant for different versions of the generic problem under consideration. Section 4 sets out a set of principles for distributional comparisons in terms of aggregate divergence and show how these characterise a class of measures. Section 5 performs a set of experiments and applications using the proposed measures and UK income data. Section 6 concludes.

2 Information and income distribution

Comparisons of distributions using information-theoretic approaches has involved comparing entropy-based measures which quantify the discrepancies between the probability distributions. This concept was first introduced by Shannon (1948) and then further developed into a relative measure of entropy by Kullback and Leibler (1951). In this section, we show that generalised entropy inequality measures are obtained by little more than a change of variables from these entropy measures. We will then use this approach to discrepancies between distributions in order to formulate appropriate inequality measures.

2.1 Entropy: basic concept

Take a variable y distributed on support Y. Although it is not necessary for much of the discussion, it is often convenient to suppose that the distribution has a well-defined density function $f(\cdot)$ so that, by definition, $\int_Y f(y)dy = 1$. Now consider the information conveyed by the observation that an event $y \in Y$ has occurred when it is known that the density function was f. Shannon (1948) suggested a simple formulation for the information function g: the information content from an observation y when the density is f is $g\left(f(y)\right) = -\log f(y)$. The *entropy* is the expected information

$$H(f) := -E\log f(y) = -\int_Y \log f(y)\ f(y)dy. \tag{1}$$

In the case of a discrete distribution, where Y is finite with index set K and the probability of event $k \in K$ occurring is p_k, the entropy will be

$$-\sum_{k\in K} p_k\ \log p_k.$$

Clearly $g(p_k)$ decreases with p_k capturing the idea that larger is the probability of event k the smaller is the information value of an observation that k has actually occurred; if event k is known to be certain ($p_k = 1$) the observation that it has occurred conveys no information and we have $g(p_k) = -\log(p_k) = 0$. It is also clear that this definition implies that if k and k' are two independent events then $g(p_k p_{k'}) = g(p_k) + g(p_{k'})$

It is not self-evident that the additivity property of independent events is essential and so it may be appropriate to take a generalisation of the Shannon (1948) approach[5] where g is any convex function with $g(1) = 0$ (Khinchin 1957). An important special case is given by $g(f) = \frac{1}{\alpha-1}[1 - f^{\alpha-1}]$ where $\alpha > 0$ is a parameter. From this we get a generalisation of (1), the α-class entropy

$$H_\alpha(f) := Eg(f(y)) = \frac{1}{\alpha - 1}\left[1 - E(f(y)^{\alpha-1})\right], \alpha > 0. \tag{2}$$

2.2 Entropy and inequality

To transfer these ideas to the analysis of income distributions it is useful to perform a transformation similar to that outlined in Theil (1967). Suppose we specialise the model of Section 2.1 to the case of univariate probability distributions: instead of $y \in Y$, with Y as general, take $x \in \mathbb{R}_+$ where x can be thought of as "income." Let the distribution function be F so that a proportion

$$q = F(x)$$

of the population has an income less than or equal to x. Given that the population size is normalised to 1, we may define the income share function $s : [0, 1] \to [0, 1]$ as

$$s(q) := \frac{F^{-1}(q)}{\int_0^1 F^{-1}(t)\, dt} = \frac{x}{\mu} \tag{3}$$

where $F^{-1}(\cdot)$ is the inverse of the function F and μ is the mean of the income distribution. One way of reading (3) is that those located in a small neighbourhood around the q-th quantile have a share $s(q)\, dq$ in total income. It is clear that the function $s(\cdot)$ has the same properties as the regular density function $f(\cdot)$:

$$s(q) \geq 0, \text{ for all } q \qquad \text{and} \qquad \int_0^1 s(q)\, dq = 1. \tag{4}$$

[5]Using l'Hôpital's rule we can see that when $\alpha = 0$ H_α takes the form (1). For discussion of H_α see Havrda and Charvat (1967), Ullah (1996).

We may thus use $s(\cdot)$ rather than $f(\cdot)$ to characterise the income distribution. Replacing f by s in (1), we obtain

$$H(s) = -\int_0^1 s(q)\log[s(q)]dq = -\int_0^\infty \frac{x}{\mu}\log\left(\frac{x}{\mu}\right) dF(x) \tag{5}$$

The Theil inequality index is defined by

$$I_1 := \int_0^\infty \frac{x}{\mu}\log\left(\frac{x}{\mu}\right) dF(x) \tag{6}$$

and thus we have $I_1 = -H(s)$. The analogy between the Shannon entropy measure (1) and the Theil inequality measure (6) is evident and requires no more than a change of variables. The transformed version due to Theil is more useful in the context of income distribution because it enables a link to be established with several classes of inequality measures. The generalised entropy inequality measure is defined by

$$I_\alpha = \int_0^\infty \frac{1}{\alpha(\alpha-1)}\left[\left[\frac{x}{\mu}\right]^\alpha - 1\right] dF(x) \tag{7}$$

and thus, replacing f by s in (2), it is clear that $I_\alpha = -\alpha^{-1}H_\alpha(s)$, $\alpha > 0$. One of the attractions of the form (7) is that the parameter α has a natural interpretation in terms of economic welfare: for $\alpha > 0$ the measure I_α is "top-sensitive" in that it gives higher importance to changes in the top of the income distribution; $\alpha < 0$ it is particularly sensitive to changes at the bottom of the distribution; Atkinson (1970)'s index of relative inequality aversion is identical to $1 - \alpha$ for $\alpha < 1$.

2.3 Divergence entropy

It is clear that there is a close analogy between the α-class of entropy measures (2) and the generalised entropy inequality measure (7). Effectively it requires little more than a change of variables. We will now develop an approach to the problem of characterising changes in distributions using a similar type of argument.

Let the *divergence* between two densities f_2 and f_1 be $\lambda := f_1/f_2$; clearly the difference in the distributions is large when λ is far from 1. Using an entropy formulation of a divergence measure, one can measure the amount of information in λ using some convex function, $g(\lambda)$, such that $g(1) = 0$. The expected information content in f_2 with respect to f_1, or the divergence of f_2 with respect to f_1, is given by

$$H(f_1, f_2) = \int_Y g\left(\frac{f_1}{f_2}\right) f_1 dy \tag{8}$$

which is nonnegative (by Jensen's inequality) and is zero if and only if $f_2 = f_1$. Corresponding to (2), we have the class of divergence measures

$$H_\alpha(f_1, f_2) = \frac{1}{\alpha - 1} \int_Y \left[1 - f_1 \left[\frac{f_1}{f_2}\right]^{\alpha-1}\right] dy, \alpha > 0 \tag{9}$$

In the case $\alpha = 1$ we obtain the Kullback and Leibler (1951) generalisation of the Shannon entropy (1)

$$H_1(f_1, f_2) = \int_Y f_1 \log\left(\frac{f_2}{f_1}\right) dy = -E_{f_1}\left(\log\frac{f_1}{f_2}\right), \tag{10}$$

known as the *relative entropy* or *divergence measure* of f_2 from f_1. When f_2 is the uniform density, (10) becomes (1).

2.4 Discrepancy and distributional change

The transformation used to derive the Theil inequality measure from the entropy measure may also be applied to the case of divergence entropy measures. Consider a pair (x, y) jointly distributed on $\mathbb{R}_+^2$: for example x and y could represent two different definitions of income. Given that the population size is normalised to 1, we may define the income share functions s_1 and $s_2 : [0,1] \to [0,1]$ as

$$s_1(q) = \frac{F_1^{-1}(q)}{\int_0^1 F_1^{-1}(t)\,dt} = \frac{x}{\mu_1} \quad \text{and} \quad s_2(q) = \frac{F_2^{-1}(q)}{\int_0^1 F_2^{-1}(t)\,dt} = \frac{y}{\mu_2} \tag{11}$$

where F_1^{-1} is the inverse of the marginal distribution of x, F_2^{-1} is the inverse of the marginal distribution of y and μ_1, μ_2 are the means of the marginal distributions of x and y.

We may now use the concept of relative entropy to characterise the transformed distribution. Instead of considering a pair of density functions f_1, f_2, we consider a pair of income-share functions s_1, s_2. Replacing f_1 and f_2 by s_1 and s_2 in (10) we obtain

$$H_1(s_1, s_2) = -\int_0^1 s_1(q) \log\left(\frac{s_2(q)}{s_1(q)}\right) dq \tag{12}$$

A normalised version of the measure of distributional change, proposed by Cowell (1980), for two n-vectors of income $\mathbf{x}$ and $\mathbf{y}$ can be written:

$$J_1(\mathbf{x}, \mathbf{y}) := \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{\mu_1} \log\left(\frac{x_i}{\mu_1} / \frac{y_i}{\mu_2}\right). \tag{13}$$

In the case of a discrete distribution with n point masses it is clear that we have $J_1(\mathbf{x}, \mathbf{y}) = -H_1(s_1, s_2)$.

Replacing f_1 and f_2 by s_1 and s_2 in equation (9), and rearranging, we obtain

$$H_\alpha(s_1, s_2) = \frac{1}{\alpha - 1} \int_0^1 \left[1 - s_1(q)^\alpha s_2(q)^{1-\alpha}\right] dq \tag{14}$$

The J class of distributional-change measure, proposed by Cowell (1980) for two n-vectors of income $\mathbf{x}$ and $\mathbf{y}$ is

$$J_\alpha(\mathbf{x}, \mathbf{y}) := \frac{1}{n\alpha(\alpha - 1)} \sum_{i=1}^{n} \left[\left[\frac{x_i}{\mu_1}\right]^\alpha \left[\frac{y_i}{\mu_2}\right]^{1-\alpha} - 1\right], \tag{15}$$

where α takes any real value; the limiting form for $\alpha = 0$ is given by

$$J_0(\mathbf{x}, \mathbf{y}) = -\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\mu_2} \log\left(\frac{x_i}{\mu_1} / \frac{y_i}{\mu_2}\right) \tag{16}$$

and for $\alpha = 1$ is given by (13); note that $J_\alpha(\mathbf{x}, \mathbf{y}) \geq 0$ for arbitrary $\mathbf{x}$ and $\mathbf{y}$.[6] The family (15) represents an aggregate measure of *discrepancy* between two distributions. Again, for a discrete distribution with n point masses, it is clear that $J_\alpha(\mathbf{x}, \mathbf{y}) = -\alpha^{-1} H_\alpha(s_1, s_2)$. The analogy between the α-class of divergence measures and the measure of discrepancy (15) is evident and requires no more than a change of variables. Once again the parameter α has the natural welfare interpretation pointed out in section 2.2.

3 Reference distributions

The analysis in section 2 provides a natural lead into a discussion of the divergence between an Empirical Distribution Function (EDF) and a theoretical reference distribution F_*. In order to do it, for fixed q values, we need to compare the corresponding q-quantiles given by the EDF and F_*.

For instance, Figure 1 presents an EDF (red dots) and a theoretical reference distribution (blue line) on a reverse graph, with the q values in the x-axis and the income quantiles in the y-axis.

[6]To see this write (15) as

$$\sum_{i=1}^{n} \frac{y_i}{n\mu_2} \left[\psi(q_i) - \psi(1)\right], \text{ where } q_i := \frac{x_i \mu_2}{y_i \mu_1}, \psi(q) := \frac{q^\alpha}{\alpha[\alpha - 1]}$$

Because ψ us a convex function we have, for any $(q_1, ..., q_n)$ and any set of non negative weights $(w_1, ..., w_n)$ that sum to 1, $\sum_{i=1}^n w_i \psi(q_i) \geq \psi\left(\sum_{i=1}^n w_i q_i\right)$. Letting $w_i = y_i / [n\mu_2]$ and using the definition of q_i we can see that $w_i q_i = x_i / [n\mu_1]$ so we have $\sum_{i=1}^n w_i \psi(q_i) \geq \psi(1)$ and the result follows.

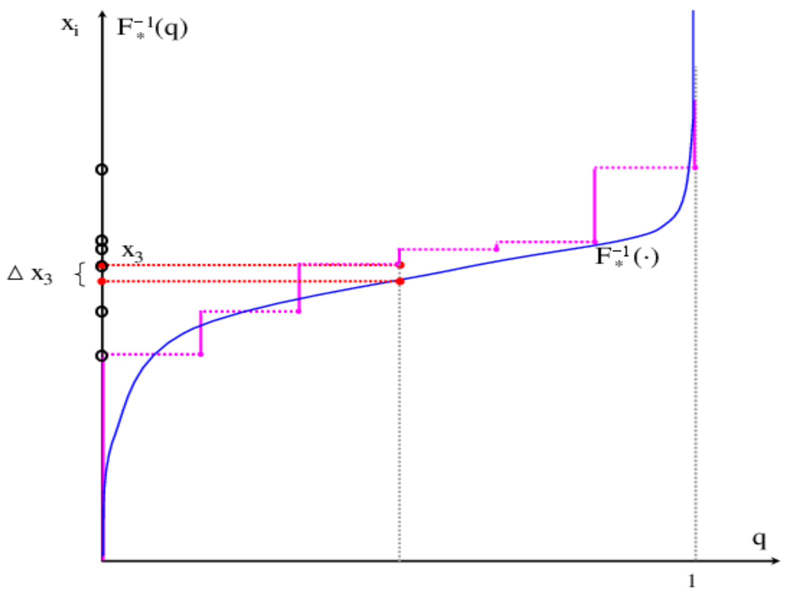

Figure 1: Quantile approach

A divergence measure between the EDF and the theoretical distribution would aggregrate discrepancies between the quantiles from the two distributions, for each values of q. In other words, we would replace x_i and y_i in (15), respectively, by $EDF^{-1}(q_i)$ and $F_{\star}^{-1}(q_i)$.

The standard approach in the statistics literature is based upon the empirical distribution function (EDF)

$$\hat{F}(x) = \frac{1}{n+1} \sum_{i=1}^{n} \iota(x_i \leq x),$$

where ι is an indicator function such that $\iota(S) = 1$ if statement S is true and $\iota(S) = 0$ otherwise.[7] Let us denote $\{x_{(1)}, x_{(2)}, ..., x_{(n)}\}$ the members of the sample in increasing order. The corresponding values given by the EDF are the adjusted sample proportion $q = \{\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1}\}$ and, for each q, the corresponding value for the reference distribution is equal to

$$y_i = F_*^{-1}\left(\frac{i}{n+1}\right) \tag{17}$$

[7]Note that we use $\frac{1}{n+1}$ rather than $\frac{1}{n}$ to avoid an obvious problem where $i = n$. Had we used $\frac{i}{n}$ in (17) then y_n would automatically be set to $\sup(X)$ where X is the support of F_*.

A divergence measure between the EDF and a theoretical reference distribution F_* would be given by replacing x_i and y_i in (15), respectively, by $EDF_*^{-1}\left(\frac{i}{n+1}\right) = x_{(i)}$ and $F_*^{-1}\left(\frac{i}{n+1}\right)$. Thus, we would have

$$J_\alpha = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left[\frac{x_{(i)}}{\hat{\mu}}\right]^{\alpha} \left[\frac{F_*^{-1}\left(\frac{i}{n+1}\right)}{\mu\left(F_*\right)}\right]^{1-\alpha} - 1 \right], \ \alpha \neq 0,1 \qquad (18)$$

The limiting forms for $\alpha = 0, 1$, defined in (16) and (13) become

$$J_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{x_{(i)}}{\hat{\mu}} \log \left(\frac{x_{(i)}}{\hat{\mu}} \Big/ \frac{F_*^{-1}\left(\frac{i}{n+1}\right)}{\mu\left(F_*\right)} \right), \qquad (19)$$

$$J_0 = -\frac{1}{n} \sum_{i=1}^{n} \frac{F_*^{-1}\left(\frac{i}{n+1}\right)}{\mu\left(F_*\right)} \log \left(\frac{x_{(i)}}{\hat{\mu}} \Big/ \frac{F_*^{-1}\left(\frac{i}{n+1}\right)}{\mu\left(F_*\right)} \right). \qquad (20)$$

This index can be used to measure the divergence between an empirical income distribution, given by a sample of individual incomes, and any theoretical reference distribution.

This index would require the choice of a specific value or values for the parameter α according to the judgment that one wants to make about the relative importance of different types of discrepancy: choosing a large positive value for α would put a lot of weight on parts of the distribution where the observed incomes x_i greatly exceed the corresponding values y_i in the reference distribution ; choosing a substantial negative value would put a lot of weight on cases where the opposite type of discrepancy arises.

3.1 The most equal reference distribution

Let us assume that the most equal income distribution is when the same amount is given to each individuals:

$$F_*^{-1}\left(\frac{i}{n+1}\right) = \hat{\mu} \qquad \text{for } i = 1, \ldots, n \qquad (21)$$

If we use this (egalitarian) distribution as the reference distribution in (18), then we have

$$J_\alpha = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left(\frac{x_i}{\hat{\mu}}\right)^{\alpha} - 1 \right], \ \alpha \neq 0,1 \qquad (22)$$

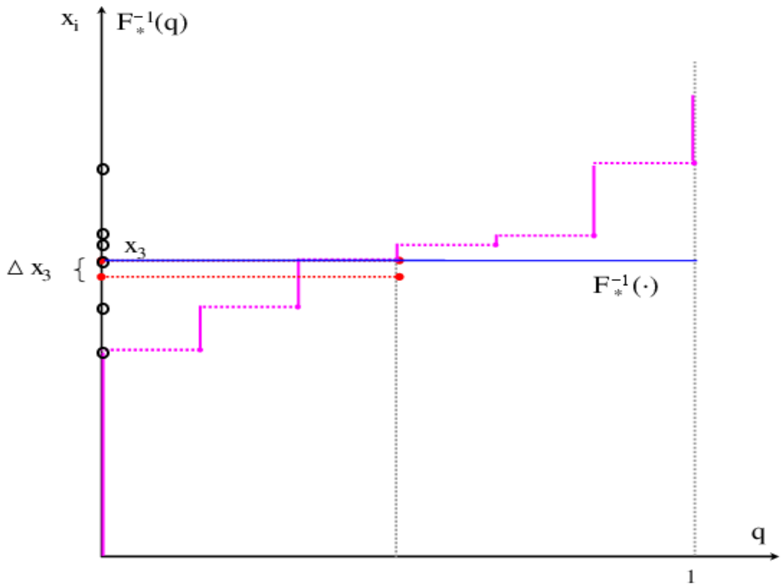

Figure 2: Quantile approach with the most equal reference distribution

and the limiting forms, for $\alpha = 0, 1$, are equal to

$$J_0 = -\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{x_i}{\hat{\mu}}\right) \quad \text{and} \quad J_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i}{\hat{\mu}} \log\left(\frac{x_i}{\hat{\mu}}\right) \tag{23}$$

These measures are nothing but the standard Generalised Entropy inequality measure. In other words, the standard GE inequality measures are divergence measures between the EDF and the most equal distribution, where everybody gets the same income. They tell us how far a distribution is from the most equal distribution. A sample with a smaller index has a more *equal* distribution.

Figure 2 presents the quantile approach for this case. We can see that the EDF is always above (below) the reference distribution for large (small) values of incomes. It makes clear that large (small) values of α would be more sensitive to changes in high (small) incomes.

3.2 The most unequal reference distribution

Rather than selecting the most equal distribution as a reference distribution, we can reverse the standard approach by using the most unequal distribution as a reference distribution. It requires one to measure how far a sample is from the most unequal distribution, rather than how far it is from the most equal distribution.

The most unequal income distribution is when one person gets all the income and the others zero:

$$F_*^{-1}\left(\frac{i}{n+1}\right) = \begin{cases} 0 & \text{for } i = 1, \ldots, n-1 \\ n\hat{\mu} & \text{for } i = n \end{cases} \tag{24}$$

If we use this distribution as the reference distribution in (18), then we have:

$$J_\alpha = \frac{1}{\alpha(\alpha-1)} \left[\left(\frac{\max x_i}{n\hat{\mu}} \right)^\alpha - 1 \right], \ \alpha < 1, \alpha \neq 0 \tag{25}$$

In the limiting case $\alpha = 0$, we have

$$J_0 = -\log\left(\frac{\max x_i}{n\hat{\mu}}\right) \tag{26}$$

This follows immediately from l'Hopital's rule. In the limiting case $\alpha = 1$, the index is undefined. This index tells us how far a distribution is from the most unequal distribution. However, two major drawbacks make this index useless in practice:

1. To be comparable for two different samples, the index should use the same reference distribution in both samples. Here $\max x_i = x_{(n)}$ is an estimate of the $n/(n+1)$-quantile in F_*. It follows that, if $n = 100$, the reference distribution is when the top 1% gets all the income, whereas if $n = 1000$ it is when the top 0.1% gets all the income. The reference distribution differs with the sample size.

2. The presence of zero incomes in the reference distribution produce undesirable properties: the index is independent on how the first $n-1$ ordered incomes are distributed. The first $n-1$ ordered incomes do not appear explicitly in the formula, the index depends on them through the mean only. It follows that, the mean being constant, the distribution of the $n-1$ first ordered incomes does not matter. For instance, the two samples $\{8,8,8,8,8,8,8,8,20\}$ and $\{1,1,1,1,15,15,15,15,20\}$ produce the same value of the index.

These two drawbacks lead us to consider the following reference distribution, where the top $100k\%$ richest gets $100p\%$ of the total income:

$$F_*^{-1}\left(\frac{i}{n+1}\right) = \begin{cases} (1-p)\hat{\mu}/(1-k) & \text{for } i = 1, \ldots, \lceil n(1-k) \rceil \\ p\hat{\mu}/k & \text{for } i = \lceil n(1-k) \rceil + 1, \ldots, n \end{cases} \tag{27}$$

with $0 \leq k \leq 1$, $0 \leq p \leq 1$ and $\lceil z \rceil$ denotes the smallest integer not less than z. Small values of k and large values of p produce very unequal distributions, where *a few* people get *nearly* all the income, and the rest get nearly zero. For instance, in setting $k = 0.01$ and $p = 0.99$ we take the case where the top 1% richest gets 99% of the total income. If we use this distribution as the reference distribution in (18), we obtain:

$$J_{\alpha,k,p} = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left(\frac{x_{(i)}}{\hat{\mu}} \right)^{\alpha} c_i^{1-\alpha} - 1 \right], \quad \alpha \neq 0, 1, \tag{28}$$

where

$$c_i = \begin{cases} (1-p)/(1-k) & \text{if } i \leq \lceil n(1-k) \rceil \\ p/k & \text{if } i > \lceil n(1-k) \rceil \end{cases} \tag{29}$$

The limiting forms for $\alpha = 0, 1$, defined in (19) and (20) become

$$J_{1,k,p} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_{(i)}}{\hat{\mu}} \log \left(\frac{x_{(i)}}{c_i \hat{\mu}} \right) \quad \text{and} \quad J_{0,k,p} = -\frac{1}{n} \sum_{i=1}^{n} c_i \log \left(\frac{x_{(i)}}{c_i \hat{\mu}} \right). \tag{30}$$

There is two interesting special cases. If $p = k$, everybody gets the same income value, $\hat{\mu}$, and the reference distribution is the most equal distribution. If $k = 1/n$ and $p = 1$, only one individual gets all the income, $n\hat{\mu}$, and the reference distribution is the most unequal distribution.

In practice, k and p have to be fixed: (1) to avoid the first drawback, k and p should be independent of the sample size, with $k > 1/n$ and $p > 1/n$; (2) to avoid the second drawback, zero incomes are not allowed in the reference distribution, that is, if $p = 1$ we have $k = 1$. Finally, to make our index $J_{\alpha,k,p}$ useful in practice, we need to use constant values, such that

$$1/n < k < 1 \quad \text{and} \quad 1/n < p < 1, \quad \text{or} \quad p = k = 1. \tag{31}$$

In empirical studies, we could use several values of k and p. For instance, $k = 1 - p = 0.05, 0.01, 0.005$, correspond to the reference distributions with the top 5%, 1% and 0.5% getting, respectively, 95%, 99% and 99.5% of the total income.

3.3 Other reference distributions

Clearly, other reference distributions could be used. For instance, if we assume that productive talents are distributed in the population according to a continuous distribution of talents F_* and that wages should be related to talent, a situation in which everyone received the same income to everybody might be considered as unfair. In this case one might use F_* as the reference distribution and make use of the index (18): any deviation from F_* would come from something else than talent. If total income is finite, it makes sense to use a distribution defined on a finite support. For instance, we could use a Uniform distribution or a Beta distribution with two parameters, which can provide a variety of appropriate shapes.

4 Axiomatic foundation

We may put the informal discussion of the use of distributional-change measures on to a rigorous footing using the representation of the problem in section 4.1 and the principles described in section 4.2.

4.1 Representation of the problem

The distributional change problem can be characterised as the relationship between two n-vectors of incomes $\mathbf{x}$ and $\mathbf{y}$. An alternative equivalent approach is to work with $\mathbf{z} := (z_1, z_2, ..., z_n)$, where each z_i is the ordered pair (x_i, y_i), $i = 1, ..., n$ and belongs to a set Z, which we will take to be a connected subset of $\mathbb{R}_+ \times \mathbb{R}_+$. The divergence issue clearly focuses on the discrepancies between the x-values and the y-values. To capture this we introduce a discrepancy function $d : Z \to \mathbb{R}$ such that $d(z_i)$ is strictly increasing in $|x_i - y_i|$. Write the vector of discrepancies as

$$\mathbf{d}(\mathbf{z}) := (d(z_1), ..., d(z_n)).$$

The problem can then be approached in two steps.

1. We represent the problem as one of characterising a weak ordering[8] $\succeq$ on
$$Z^n := \underbrace{Z \times Z \times ... \times Z}_{n}.$$
where, for any $\mathbf{z}, \mathbf{z}' \in Z^n$ the statement "$\mathbf{z} \succeq \mathbf{z}'$" should be read as "the income pairs in $\mathbf{z}$ constitute at least as close according to $\succeq$ as the income pairs in $\mathbf{z}'$." From $\succeq$ we may derive the antisymmetric part $\succ$ and symmetric part $\sim$ of the ordering.[9]

[8]This implies that it has the minimal properties of completeness, reflexivity and transitivity.

[9]For any $\mathbf{z}, \mathbf{z}' \in Z^n$ "$\mathbf{z} \succ \mathbf{z}'$" means "$[\mathbf{z} \succeq \mathbf{z}'] \,\&\, [\mathbf{z}' \not\succeq \mathbf{z}]$"; "$\mathbf{z} \sim \mathbf{z}'$" means "$[\mathbf{z} \succeq \mathbf{z}'] \,\&\, [\mathbf{z}' \succeq \mathbf{z}]$".

2. We use the function representing $\succeq$ to generate the index J.

In the first stage of step 1 we introduce some properties for $\succeq$, many of which are standard in choice theory and welfare economics.[10]

4.2 Basic structure

Axiom 1 *[Continuity] $\succeq$ is continuous on Z^n.*

Axiom 2 *[Monotonicity] If $\mathbf{z}, \mathbf{z}' \in Z^n$ differ only in their ith component then $d(x_i, y_i) < d(x_i', y_i') \iff \mathbf{z} \succ \mathbf{z}'$.*

Axiom 3 *[Symmetry] For any $\mathbf{z}, \mathbf{z}' \in Z^n$ such that $\mathbf{z}'$ is obtained by permuting the components of $\mathbf{z}$: $\mathbf{z} \sim \mathbf{z}'$.*

In view of Axiom 3 we may without loss of generality impose a simultaneous ordering on the x and y components of $\mathbf{z}$, for example $x_1 \leq x_2 \leq ... \leq x_n$ and $y_1 \leq y_2 \leq ... \leq y_n$.[11] For any $\mathbf{z} \in Z^n$ denote by $\mathbf{z}(\zeta, i)$ the member of Z^n formed by replacing the ith component of $\mathbf{z}$ by $\zeta \in Z$.

Axiom 4 *[Independence] For $\mathbf{z}, \mathbf{z}' \in Z^n$ such that: $\mathbf{z} \sim \mathbf{z}'$ and $z_i = z_i'$ for some i then $\mathbf{z}(\zeta, i) \sim \mathbf{z}'(\zeta, i)$ for all $\zeta \in [z_{i-1}, z_{i+1}] \cap [z_{i-1}', z_{i+1}']$.*

If $\mathbf{z}$ and $\mathbf{z}'$ are equivalent in terms of overall discrepancy and the discrepancy at position i is the same in the two cases then a local variation at i simultaneously in $\mathbf{z}$ and $\mathbf{z}'$ has no overall effect.

Axiom 5 *[Zero local discrepancy] Let $\mathbf{z}, \mathbf{z}' \in Z^n$ be such that, for some i and j, $x_i = y_i$, $x_j = y_j$, $x_i' = x_i + \delta$, $y_i' = y_i + \delta$, $x_j' = x_j - \delta$, $y_j' = y_j - \delta$ and, for all $k \neq i, j$, $x_k' = x_k$, $y_k' = y_k$. Then $\mathbf{z} \sim \mathbf{z}'$.*

The principle states that if there is zero local discrepancy at two positions in the distribution then moving x-income and y-income simultaneously from one position to the other has no effect on the overall discrepancy.

[10] Note that the derivation which follows differs from that provided in Cowell (1985) used to establish the class of measures of distributional change using explicit assumptions of differentiability and additive separability. Here we adopt a minimalist approach that uses neither of these strong assumptions and that focuses directly on the divergence issue.

[11] In the general distributional change problem $\mathbf{x}$ and $\mathbf{y}$ could be arbitrary vectors but in the present case, of course, the components of $\mathbf{x}$ and $\mathbf{y}$ will be in the same order.

Theorem 1 *Given Axioms 1 to 5 (a) $\succeq$ is representable by the continuous function given by*

$$\sum_{i=1}^{n} \phi_i(z_i), \forall \mathbf{z} \in Z^n \tag{32}$$

where, for each i, $\phi_i : Z \to \mathbb{R}$ is a continuous function that is strictly decreasing in $|x_i - y_i|$ and (b)

$$\phi_i(x, x) = a_i + b_i x \tag{33}$$

Proof. Axioms 1 to 5 imply that $\succeq$ can be represented by a continuous function $\Phi : Z^n \to \mathbb{R}$ that is increasing in $|x_i - y_i|$, $i = 1, ..., n$. Using Axiom 4 part (a) of the result follows from Theorem 5.3 of Fishburn (1970). Now take $\mathbf{z}'$ and $\mathbf{z}$ in as specified in Axiom 5. Using (32) and it is clear that $\mathbf{z} \sim \mathbf{z}'$ if and only if

$$\phi_i(x_i + \delta, x_i + \delta) - \phi_i(x_i, x_i) - \phi_j(x_j + \delta, x_j + \delta) + \phi_j(x_j + \delta, x_j + \delta) = 0$$

which can only be true if

$$\phi_i(x_i + \delta, x_i + \delta) - \phi_i(x_i, x_i) = f(\delta)$$

for arbitrary x_i and δ. This is a standard Pexider equation and its solution implies (33). ■

Corollary 1 *Since $\succeq$ is an ordering it is also representable by*

$$\phi\left(\sum_{i=1}^{n} \phi_i(z_i)\right) \tag{34}$$

where, ϕ_i is defined as in (32), (33). and $\phi : \mathbb{R} \to \mathbb{R}$ continuous and strictly monotonic increasing.

This additive structure means that we can proceed to evaluate overall discrepancy one income-position at a time. The following axiom imposes a very weak structural requirement, namely that the ordering remains unchanged by some uniform scale change to both x-values and y-values simultaneously. As Theorem 2 shows it is enough to induce a rather specific structure on the function representing $\succeq$.

Axiom 6 *[Income scale irrelevance] For any $\mathbf{z}, \mathbf{z}' \in Z^n$ such that $\mathbf{z} \sim \mathbf{z}'$, $t\mathbf{z} \sim t\mathbf{z}'$for all $t > 0$.*

Theorem 2 *Given Axioms 1 to 6* $\succeq$ *is representable by*

$$\phi\left(\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right)\right) \tag{35}$$

where h_i *is a real-valued function.*

Proof. Using the function Φ introduced in the proof of Theorem 1 Axiom 6 implies

$$\begin{aligned}\Phi(\mathbf{z}) &= \Phi(\mathbf{z}') \\ \Phi(t\mathbf{z}) &= \Phi(t\mathbf{z}')\end{aligned}$$

and so, since this has to be true for arbitrary $\mathbf{z}, \mathbf{z}'$ we have

$$\frac{\Phi(t\mathbf{z})}{\Phi(\mathbf{z})} = \frac{\Phi(t\mathbf{z}')}{\Phi(\mathbf{z}')} = \psi(t)$$

where ψ is a continuous function $\mathbb{R} \to \mathbb{R}$. Hence, using the ϕ_i given in (32), we have for all :

$$\phi_i(tz_i) = \psi(t)\,\phi_i(z_i)\; i = 1, ..., n.$$

or, equivalently

$$\phi_i(tx_i, ty_i) = \psi(t)\,\phi_i(x_i, y_i)\,, i = 1, ..., n. \tag{36}$$

So, in view of Aczél and Dhombres (1989), page 346 there must exist $c \in \mathbb{R}$ and a function $h_i : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\phi_i(x_i, y_i) = x_i^c h_i\left(\frac{x_i}{y_i}\right). \tag{37}$$

From (33) and (37) it is clear that

$$\phi_i(x_i, x_i) = x_i^c h_i(1) = a_i + b_i x_i, \tag{38}$$

which, if$\phi_i(x, x)$ is non-constant in x, implies $c = 1$. Putting (37) with $c = 1$ into (34) gives the result. ■

This result is important but limited since the function h_i is essentially arbitrary: we need to impose more structure.

4.3 Income discrepancy

We now focus on the way in which one compares the (x, y) discrepancies in different parts of the income distribution. The form of (35) suggests that discrepancy should be characterised terms of proportional differences:

$$d(z_i) = \max\left(\frac{x_i}{y_i}, \frac{y_i}{x_i}\right).$$

This is the form for d that we will assume from this point onwards. We also introduce:

Axiom 7 *[Discrepancy scale irrelevance] Suppose there are* $\mathbf{z}_0, \mathbf{z}_0' \in Z^n$ *such that* $\mathbf{z}_0 \sim \mathbf{z}_0'$. *Then for all* $t > 0$ *and* $\mathbf{z}, \mathbf{z}'$ *such that* $d(\mathbf{z}) = td(\mathbf{z}_0)$ *and* $d(\mathbf{z}') = td(\mathbf{z}_0')$*:* $\mathbf{z} \sim \mathbf{z}'$.

The principle states this. Suppose we have two discrepancy profiles $\mathbf{z}_0$ and $\mathbf{z}_0'$ that are regarded as equivalent under $\succeq$. Then scale up (or down) all the income discrepancies in $\mathbf{z}_0$ and $\mathbf{z}_0'$ by the same factor t. The resulting pair of discrepancy profiles $\mathbf{z}$ and $\mathbf{z}'$ will also be equivalent.[12]

Theorem 3 *Given Axioms 1 to 7* $\succeq$ *is representable by*

$$\phi\left(\sum_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha}\right) \tag{39}$$

where $\alpha \neq 1$ *is a constant.*[13]

Proof. Take the special case where, in distribution $\mathbf{z}_0'$ the income discrepancy takes the same value r at all n income positions. If (x_i, y_i) represents a typical component in $\mathbf{z}_0$ then $\mathbf{z}_0 \sim \mathbf{z}_0'$ implies

$$r = \psi\left(\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right)\right) \tag{40}$$

where ψ is the solution in r to

$$\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right) = \sum_{i=1}^{n} x_i h_i(r) \tag{41}$$

[12]Also note that Axiom 7 can be stated equivalently by requiring that, for a given $\mathbf{z}_0, \mathbf{z}_0' \in Z^n$ such that $\mathbf{z}_0 \sim \mathbf{z}_0'$, either (a) any $\mathbf{z}$ and $\mathbf{z}'$ found by rescaling the x-components will be equivalent or (b) any $\mathbf{z}$ and $\mathbf{z}'$ found by rescaling the y-components will be equivalent.

[13]The following proof draws on Ebert (1988).

In (41) can take the x_i as fixed weights. Using Axiom 7 in (40) requires

$$tr = \psi\left(\sum_{i=1}^{n} x_i h_i\left(t\frac{x_i}{y_i}\right)\right), \text{ for all } t > 0. \tag{42}$$

Using (41) we have

$$\sum_{i=1}^{n} x_i h_i\left(t\psi\left(\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right)\right)\right) = \sum_{i=1}^{n} x_i h_i\left(t\frac{x_i}{y_i}\right) \tag{43}$$

Introduce the following change of variables

$$u_i := x_i h_i\left(\frac{x_i}{y_i}\right), i = 1, ..., n \tag{44}$$

and write the inverse of this relationship as

$$\frac{x_i}{y_i} = \psi_i\left(u_i\right), i = 1, ..., n \tag{45}$$

Substituting (44) and (45) into (43) we get

$$\sum_{i=1}^{n} x_i h_i\left(t\psi\left(\sum_{i=1}^{n} u_i\right)\right) = \sum_{i=1}^{n} x_i h_i\left(t\psi_i\left(u_i\right)\right). \tag{46}$$

Also define the following functions

$$\theta_0\left(u, t\right) := \sum_{i=1}^{n} x_i h_i\left(t\psi\left(u\right)\right) \tag{47}$$

$$\theta_i\left(u, t\right) := x_i h_i\left(t\psi_i\left(u\right)\right), i = 1, ..., n. \tag{48}$$

Substituting (47),(48) into (46) we get the Pexider functional equation

$$\theta_0\left(\sum_{i=1}^{n} u_i, t\right) = \sum_{i=1}^{n} \theta_i\left(u_i, t\right)$$

which has as a solution

$$\theta_i\left(u, t\right) = b_i\left(t\right) + B\left(t\right) u, i = 0, 1, ..., n$$

where

$$b_0\left(t\right) = \sum_{i=1}^{n} b_i\left(t\right)$$

– see Aczél (1966), page 142. Therefore we have

$$h_i\left(t\frac{x_i}{y_i}\right) = \frac{b_i(t)}{x_i} + B(t)\,h_i\left(\frac{x_i}{y_i}\right), i = 1, ..., n \tag{49}$$

From Eichhorn (1978), Theorem 2.7.3 the solution to (49) is of the form

$$h_i(v) = \begin{matrix} \beta_i v^{\alpha-1} + \gamma_i, & \alpha \neq 1 \\ \beta_i \log v + \gamma_i & \alpha = 1 \end{matrix} \tag{50}$$

where $\beta_i > 0$ is an arbitrary positive number. Substituting for $h_i(\cdot)$ from (50) into (2) for the case where β_i is the same for all i gives the result. ■

4.4 The J index

For the required index use the "natural" cardinalisation of the function (39), $\sum_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha}$, and normalise with reference to the case where both the observed and the modelled distribution exhibit complete equality, so $x_i = \mu_1$ and $y_i = \mu_2$ for all i. This gives the following class of measures of *divergence* (aggregated discrepancy):

$$J_\alpha(\mathbf{x}, \mathbf{y}) := \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left[\frac{x_i}{\mu_1}\right]^{\alpha} \left[\frac{y_i}{\mu_2}\right]^{1-\alpha} - 1 \right]. \tag{51}$$

This normalised index can be implemented straightforwardly for a proposed model of an empirical distribution.[14] Of course this would require the choice of a specific value or values for the parameter α in (51).[15]

5 Implementation

We now look at the practicalities of the class of measures J_α defined in (18) and $J_{\alpha,k,p}$ defined in (28).

5.1 Statistical properties

It is necessary to establish the existence of an asymptotic distribution for J_α and $J_{\alpha,k,p}$ in order to justify its use in practice. If the most equal distribution is taken as the reference distribution ($k = p = 1$), the index $J_{\alpha,1,1}$ is nothing

[14] The form (51) implies that it is valid for mean-normalised distributions.

[15] Compare this with the discussion of the interpretation of α in terms of upper- and lower-tail sensitivity in the context of inequality (page 4).

but the standard GE inequality measure, which is asymptotically Normal and has well-known statistical properties.[16] If a continuous distribution is taken as the reference distribution, it can be shown that the limiting distribution of nJ_α is that of

$$\frac{1}{2\mu_{F_*}}\left[\int_0^1 \frac{B^2(t)\mathrm{d}t}{F_*^{-1}(t)f_*^2(F_*^{-1}(t))} - \frac{1}{\mu_{F_*}}\left(\int_0^1 \frac{B(t)\mathrm{d}t}{f_*(F_*^{-1}(t))}\right)^2\right] \tag{52}$$

where f_* is the density of distribution F_* and $B(t)$ is a Brownian bridge. This random variable can have an infinite expectation. It is only if F_* has a bounded support that the limiting distribution has reasonable properties – see Cowell et al. (2011) and Davidson (2011) for more details. If we use a continuous parametric reference distribution, since total income is finite, it makes sense to use a distribution F_* defined on a bounded support only. For instance, one could use a Uniform distribution or a Beta distribution with two parameters, which can provide many different shapes. The same approach can be used for $nJ_{\alpha,k,p}$, noting that the last statistic is equivalent to the statistic defined in (9) in Cowell et al. (2011), where $2i/(n+1)$ is replaced by c_i defined in (29).

In the two cases, the limiting distribution of J_α and $J_{\alpha,k,p}$ exists, but is not tractable. It is enough to justify the use of bootstrap methods for making inference. To compute a bootstrap confidence interval, we generate B bootstrap samples by resampling from the original data, and then, for each resample, we compute the index J. We obtain B bootstrap statistics, J_α^b, $b = 1, \ldots, B$. The percentile bootstrap confidence interval is equal to

$$CI_{perc} = [c^b_{0.025}\,;\; c^b_{0.975}] \tag{53}$$

where $c^b_{0.025}$ and $c^b_{0.975}$ are the 2.5 and 97.5 percentiles of the EDF of the bootstrap statistics - for a comprehensive discussion on bootstrap methods, see Davison and Hinkley (1997), Davidson and MacKinnon (2006). For well-known reasons – see Davison and Hinkley (1997) or Davidson and MacKinnon (2000) – the number B should be chosen so that $(B+1)/100$ is an integer: here we set $B = 999$ unless otherwise stated.

To be used in practice, we need to determine the finite sample properties of J_α and $J_{\alpha,k,p}$. 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 generate a coverage error rate close to

[16] Among others, see Cowell and Flachaire (2007), Davidson and Flachaire (2007), Schluter and van Garderen (2009), Schluter (2011), Davidson (2011)

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 bootstrap confidence intervals in several experimental designs.

In our experiments, samples are drawn from a lognormal distribution. For fixed values of α, k, p and n, we draw 10 000 samples. For each sample we compute J_α or $J_{\alpha,k,p}$ 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.[17] Confidence intervals perform well in finite samples if the coverage error rate is close to the nominal value, that is, close to 0.05.

Table 1 presents the coverage error rate of bootstrap confidence intervals at 95% of J_α and $J_{\alpha,k,p}$ for several reference distributions. The standard GE measures use the most equal reference distribution, it corresponds to $J_{\alpha,1,1}$. When "the top 1% richest gets 99% of the income" is the reference distribution we use the index $J_{\alpha,0.01,0.99}$; when "the top 5% richest gets 99% of the income" is the reference distribution we use $J_{\alpha,0.05,0.99}$. In addition, we examine J_α with two continuous (bounded) parametric reference distributions, the Beta(1,1) distribution which is equal to the Uniform(0,1), and the Beta(2,2) which is a symmetric inverted-U-shape distribution. Table 1 shows that the finite sample properties of the indices with alternative reference distributions are not very different from those of the standard GE measures, except for $J_{\alpha,0.01,0.99}$ when $n \leq 500$. The coverage error rate is close to 0.05 for very large samples. For small and moderate samples, further investigations are required to improve the finite sample properties, with, for instance, a fast double or triple bootstrap, see Davidson and MacKinnon (2007) and Davidson and Trokic (2011).[18]

5.2 Application

Let us compare the performance of the statistic $J_{\alpha,k,p}$ with that of conventional GE inequality measures using as a case study UK income data from the Family Expenditure Survey, for years 1979 and 1988.[19] In Table 5.2

[17]The true values are computed replacing $x_{(i)}$ in (18) by $F^{-1}(\frac{i}{n+1})$, where F is the distribution of x, that is, the lognormal distribution in our experiments.

[18]Such developments are beyond the scope of this paper and are the subject of future research.

[19]The application uses the "before housing costs" income variable of the Family Expenditure Survey for years 1979 and 1988 (Department of Work and Pensions 2006), deflated and equivalised using the McClement's adult-equivalence scale, excluding households with self-employed individuals. We exclude households with self-employed individuals as re-

we present the results of indices J_α and $J_{\alpha,k,p}$ estimated with three different types of reference distribution, along with bootstrap confidence intervals at 95%.

Equality

The top panel of Table 5.2 presents estimates of $J_{\alpha,k,p}$ using an "equality" reference distribution. Clearly, when we select the most equal distribution as the reference distribution, i.e. $k = p = 1$, the index $J_{\alpha,k,p}$ is reduced to the standard GE inequality measure. Estimates for standard GE measures, $J_{\alpha,1,1}$ are tabulated in the first row, for values of α ranging from -1 to 2.[20] When $\alpha = 1$, $J_{\alpha,1,1}$ is the Theil index. For values of $\alpha = 0.5, 1, 1.5, 2$, $J_{\alpha,1,1}$ represents the several (transformed) Atkinson indices.[21] All estimates of standard GE measures increase considerably between 1979 and 1988, suggesting a significant rise in inequality in the 80s.

Extreme inequality

The key point highlighted in earlier sections was that changing the reference distribution from which we measure the distance of the empirical distribution opens up the possibility for researchers to choose the exact distribution from which they wish to measure distance of the empirical distribution. While standard GE indices tell us about the distance of the empirical distribution from an equal reference distribution, one can change the focus to that of its distance from an *unequal* reference distribution. In the second panel of Table 5.2 we present estimates of $J_{\alpha,k,p}$ using several "extreme inequality" reference distributions. The interpretation of the size of the $J_{\alpha,k,p}$ index now is the reverse of the interpretation of standard GE measures. For a standard GE inequality measure, a small value of $J_{\alpha,1,1}$ corresponds to the empirical distribution being close to the equal reference distribution compared to that of a large value of $J_{\alpha,1,1}$. However, for an unequal reference distribution a small value of $J_{\alpha,k,p}$ corresponds to the empirical distribution being close to the particular "extreme inequality" reference distribution that has been specified.

To illustrate we focus on two different unequal reference distributions: one, where the top 1% of the income distribution receive 99% of the income,

---

ported incomes are known to be misrepresented. The years 1979 and 1988 have been chosen to represent the maximum recorded difference in inequality across the available years, post-1975.

[20] A large value of α implies greater weight on parts of the distribution where the observed incomes are vastly different from the corresponding values in the reference distribution.

[21] See Cowell (2011) for details.

and second, where the top 5% of the income distribution receive 99% of the income. From Table 5.2 we can see that, with one exception, the values of $J_{\alpha,k,p}$ have dropped between years 1979 and 1988: in other words, it is almost always true that the distance from the "extreme inequality" reference distribution has decreased. The exception is the case $(k = 0.05, p = 0.99, \alpha = 2)$ where the movement relative to the reference distribution is not significant. The implication is that UK inequality grew during the 1980s whether one interprets this in terms of distance from equality, or as distance from a reference unequal distribution, except for one case. This case concerns top-sensitive inequality where, in terms of "distance from maximum inequality," the change in the distribution is inconclusive.

Theoretical distribution

Finally let us consider how inequality changed using a continuous reference distribution F_*. The last panel of Table 5.2, tabulates the results for three such F_* (introduced in Section 5.1) taken from the Beta distribution family. Did UK income inequality, interpreted as a distance from a Beta-family reference distribution increase? We can see that the values of J_α are not statistically different between 1979 and 1988 when the Beta(1,1) (uniform) or Beta(2,5) (unimodal, right skewed) is used as the reference distribution distribution, while they are statistically different when the Beta$(2, 2)$ (unimodal symmetric) is used as the reference distribution.

The estimates of the standard GE inequality measures $J_{\alpha,1,1}$ and of those of $J_{\alpha,k,p}$ and J_α in Table 5.2 provide us with different information about divergence of the empirical distribution from the chosen reference distribution. By varying the values of k and p, one can specify the exact skewness of the reference distribution one would like to measure distance of the empirical distribution from. Likewise, by varying the values of α one can focus on different parts of the income distribution. A large value of α implies a greater weight on parts of the distribution where the observed incomes are vastly different from the corresponding values in the reference distribution. Finally, one can choose specific parametric distributions which correspond to the relevant reference distribution that the researcher is interested in.

6 Conclusion

The problem of comparing pairs of distributions is a widespread one in distributional analysis. It is often treated on an ad-hoc basis by invoking the concept of norm incomes and an arbitrary inequality index.

Our approach to the issue is a natural generalisation of the concept of inequality indices where the implicit reference distribution is the trivial perfect-equality distribution. It is also a natural application of information theory to assessment of income distributions. The approach uses the same ingredients as loss functions applied in other economic contexts. Its intuitive appeal is supported by the type of axiomatisation that is common in modern approaches to inequality measurement and other welfare criteria. The axiomatisation yields indices that can be interpreted as measures of discrepancy. They are related to the concept of divergence entropy in the context of information theory. Furthermore, they offer a degree of control to the researcher in that the J_α indices form a class of measures that can be calibrated to suit the nature of the economic problem under consideration. Members of the class have a distributional interpretation that is close to members of the well-known generalised-entropy class of inequality indices.

In effect the user of the J_α-index is presented with two key questions:

1. the income discrepancies underlying inequality are with reference to what?

2. to what kind of discrepancies do you want the measure to be particularly sensitive?

As our empirical illustration has shown, different responses to these two key questions provide different interpretations from the same set of facts.

References

Aczél, J. (1966). *Lectures on Functional Equations and their Applications.* Number 9 in Mathematics in Science and Engineering. New York: Academic Press.

Aczél, J. and J. G. Dhombres (1989). *Functional Equations in Several Variables.* Cambridge: Cambridge University Press.

Almås, I., A. W. Cappelen, J. T. Lind, E. O. Sorensen, and B. Tungodden (2011). Measuring unfair (in)equality. *Journal of Public Economics 95*(7-8), 488–499.

Almås, I., T. Havnes, and M. Mogstad (2011). Baby booming inequality? demographic change and earnings inequality in Norway, 1967-2000. *Journal Of Economic Inequality 9*, 629–650.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory 2*, 244–263.

Bartels, C. P. A. (1977). *Economic Aspects of Regional Welfare, Income Distribution and Unemployment*, Volume 9 of *Studies in applied regional science*. Leiden: Martinus Nijhoff Social Sciences Division.

Cowell, F. A. (1980). Generalized entropy and the measurement of distributional change. *European Economic Review 13*, 147–159.

Cowell, F. A. (1985). Measures of distributional change: An axiomatic approach. *Review of Economic Studies 52*, 135–151.

Cowell, F. A. (2011). *Measuring Inequality* (Third ed.). Oxford: Oxford University Press.

Cowell, F. A., R. Davidson, and E. Flachaire (2011). Goodness of fit: an axiomatic approach. Working Paper 2011/50, Greqam.

Cowell, F. A. and E. Flachaire (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics 141*, 1044–1072.

Davidson, R. (2011). Statistical inference in the presence of heavy tails. *Econometrics Journal*, forthcoming.

Davidson, R. and E. Flachaire (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics 141*, 141 – 166.

Davidson, R. and J. MacKinnon (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics and Data Analysis 51*, 3259–3281.

Davidson, R. and J. G. MacKinnon (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews 19*, 55–68.

Davidson, R. and J. G. MacKinnon (2006). Bootstrap methods in econometrics. In T. C. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics*, Volume 1 Econometric Theory, Chapter 23. London: Palgrave- Macmillan.

Davidson, R. and M. Trokic (2011). The iterated bootstrap. Paper presented at the Third French Econometrics Conference, Aix-en-Provence.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods*. Cambridge: Cambridge University Press.

Department of Work and Pensions (2006). *Households Below Average Income 1994/95-2004/05*. London: TSO.

Devooght, K. (2008). To each the same and to each his own: A proposal to measure responsibility-sensitive income inequality. *Economica 75*, 280–295.

Ebert, U. (1984). Measures of distance between income distributions. *Journal of Economic Theory 32*, 266–274.

Ebert, U. (1988). Measurement of inequality: an attempt at unification and generalization. *Social Choice and Welfare 5*, 147–169.

Eichhorn, W. (1978). *Functional Equations in Economics*. Reading Massachusetts: Addison Wesley.

Fishburn, P. C. (1970). *Utility Theory for Decision Making*. New York: John Wiley.

Garvy, G. (1952). Inequality of income: Causes and measurement. In *in Eight Papers on Size Distribution of Income*, Volume 15. New York: National Bureau of Economic Research.

Havrda, J. and F. Charvat (1967). Quantification method in classification processes: concept of structural α-entropy. *Kybernetica 3*, 30–35.

Jenkins, S. P. and M. O'Higgins (1989). Inequality measurement using norm incomes - were Garvy and Paglin onto something after all? *Review of Income and Wealth 35*, 245–282.

Khinchin, A. I. (1957). *Mathematical Foundations of Information Theory*. New York: Dover.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics 22*, 79–86.

Magdalou, B. and R. Nock (2011). Income distributions and decomposable divergence measures. *Journal of Economic Theory 146*, 2440–2454.

Nygård, F. and A. Sandström (1981). *Measuring Income Inequality.* Stockholm, Sweden: Almquist Wicksell International.

Paglin, M. (1975). The measurement and trend of inequality: a basic revision. *American Economic Review 65*, 598–609.

Schluter, C. (2011). The econometrics of inequality measurement. *Econometrics Journal*, forthcoming.

Schluter, C. and K. van Garderen (2009). Edgeworth expansions and normalizing transforms for inequality measures. *Journal of Econometrics 150*, 16–29.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal 106*, 379–423 and 623–656.

Theil, H. (1967). *Economics and Information Theory.* Amsterdam: North Holland.

Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *Journal of Statistical Planning and Inference 49*, 137–162.

Wertz, K. (1979). The measurement of inequality: comment. *American Economic Review 79*, 670–72.

α	-1	0	0.5	1	2
Equal Reference Distribution					
Standard GE measures (k=p=1)					
$n = 100$	0.0753	0.0734	0.0832	0.0912	0.1166
$n = 200$	0.0747	0.0667	0.0713	0.0785	0.1024
$n = 500$	0.0669	0.0673	0.0716	0.0781	0.0983
$n = 1000$	0.0658	0.0606	0.0642	0.0709	0.0878
$n = 2000$	0.0567	0.0565	0.0620	0.0658	0.0831
$n = 5000$	0.0557	0.0562	0.0606	0.0672	0.0809
Unequal Reference Distributions					
Top 5% gets 99% of the income (k=0.05, p=0.99)					
$n = 100$	0.0722	0.0887	0.0983	0.1005	0.0395
$n = 200$	0.0597	0.0662	0.0733	0.0813	0.0482
$n = 500$	0.0594	0.0577	0.0638	0.0681	0.0584
$n = 1000$	0.0553	0.0543	0.0572	0.0619	0.0575
$n = 2000$	0.0581	0.0557	0.0552	0.0590	0.0564
$n = 5000$	0.0526	0.0531	0.0562	0.0588	0.0569
Top 1% gets 99% of the income (k=0.01, p=0.99)					
$n = 100$	0.2220	0.2221	0.2172	0.1347	0.0325
$n = 200$	0.1601	0.1689	0.1705	0.1265	0.0275
$n = 500$	0.0998	0.1117	0.1188	0.1064	0.0346
$n = 1000$	0.0703	0.0788	0.0867	0.0878	0.0422
$n = 2000$	0.0581	0.0642	0.0682	0.0717	0.0486
$n = 5000$	0.0558	0.0598	0.0616	0.0627	0.0504
Continuous Reference Distributions					
Beta(1,1)					
$n = 100$	0.0830	0.0877	0.0923	0.0981	0.1162
$n = 200$	0.0703	0.0756	0.0805	0.0865	0.1029
$n = 500$	0.0689	0.0740	0.0778	0.0847	0.1011
$n = 1000$	0.0650	0.0674	0.0710	0.0766	0.0905
$n = 2000$	0.0605	0.0632	0.0645	0.0700	0.0838
$n = 5000$	0.0623	0.0638	0.0660	0.0715	0.0824
Beta(2,2)					
$n = 100$	0.0778	0.0841	0.0896	0.0945	0.1122
$n = 200$	0.0680	0.0730	0.0764	0.0832	0.1002
$n = 500$	0.0694	0.0722	0.0762	0.0829	0.0988
$n = 1000$	0.0611	0.0656	0.0682	0.0742	0.0885
$n = 2000$	0.0574	0.0626	0.0636	0.0679	0.0834
$n = 5000$	0.0584	0.0632	0.0651	0.0694	0.0816

Table 1: Coverage error rate of bootstrap confidence intervals at 95% of J_α and $J_{\alpha,k,p}$, 10,000 replications, 499 bootstraps, and $x \sim Lognormal(0,1)$.

α	-1	0	0.5	1	2
Equal Reference Distribution					
Standard GE measures ($k = p = 1$)					
1979	0.1218	0.1056	0.1046	0.1066	0.1201
	[0.1119;0.1355]	*[0.1016;0.1097]*	*[0.1005;0.1086]*	*[0.1023;0.1111]*	*[0.1132;0.1271]*
1988	0.1836	0.1541	0.1543	0.1618	0.2096
	[0.1685;0.2018]	*[0.1468;0.1613]*	*[0.1460;0.1634]*	*[0.1508;0.1728]*	*[0.1843;0.2381]*
Unequal Reference Distributions					
Top 1% gets 99% of the income ($k = 0.01, p = 0.99$)					
1979	15.29	3.370	2.906	4.403	55.39
	[14.46;16.21]	*[3.315;3.427]*	*[2.887;2.926]*	*[4.390;4.419]*	*[55.05;55.75]*
1988	11.70	3.086	2.795	4.341	57.97
	[10.59;12.792]	*[2.982;3.182]*	*[2.749;2.836]*	*[4.300;4.378]*	*[57.32;58.66]*
Top 5% gets 99% of the income ($k = 0.05, p = 0.99$)					
1979	3.803	2.080	2.271	3.768	44.43
	[3.708;3.907]	*[2.057;2.106]*	*[2.254;2.288]*	*[3.747;3.789]*	*[44.08;44.74]*
1988	3.194	1.915	2.151	3.631	44.14
	[3.088;3.293]	*[1.882;1.945]*	*[2.123;2.175]*	*[3.591;3.665]*	*[43.47;44.73]*
Continuous Reference Distributions					
Beta(1,1) or Uniform(0,1)					
1979	0.0320	0.0406	0.0483	0.0613	0.1457
	[0.0308;0.0333]	*[0.0391;0.0421]*	*[0.0465;0.0501]*	*[0.0589;0.0638]*	*[0.1311;0.1648]*
1988	0.0339	0.0418	0.0486	0.0591	0.1125
	[0.0313;0.0373]	*[0.0383;0.0461]*	*[0.0444;0.0536]*	*[0.0538;0.0655]*	*[0.1014;0.1276]*
Beta(2,2)					
1979	0.0115	0.0132	0.0143	0.0158	0.0199
	[0.0105;0.0127]	*[0.0121;0.0146]*	*[0.0131;0.0158]*	*[0.0144;0.0175]*	*[0.0181;0.0222]*
1988	0.0210	0.0243	0.0267	0.0299	0.0405
	[0.0180;0.0242]	*[0.0204;0.0283]*	*[0.0221;0.0316]*	*[0.0242;0.0362]*	*[0.0300;0.0524]*
Beta(2,5)					
1979	0.0116	0.0138	0.0153	0.0173	0.0237
	[0.0109;0.0124]	*[0.0129;0.0147]*	*[0.0143;0.0163]*	*[0.0162;0.0185]*	*[0.0219;0.0256]*
1988	0.0121	0.0142	0.0157	0.0175	0.0231
	[0.0100;0.0145]	*[0.0116;0.0172]*	*[0.0127;0.0192]*	*[0.0141;0.0217]*	*[0.0177;0.0294]*

Table 2: Inequality indices $J_{\alpha,k,p}$ and J_α computed with different reference distributions. Data are from the Family Expenditures Surveys in UK. Bootstrap confidence intervals at 95% are given in brackets.