ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

GOODNESS-OF-FIT: AN ECONOMIC APPROACH

Frank A. Cowell, Emmanual Flachaire and Sanghamitra Bandopadhyay

Number 444
August 2009

Manor Road Building, Oxford OX1 3UQ

Goodness-of-Fit: An Economic Approach

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

April 2009

[1]London School of Economics, Sticerd, Houghton Street London WC2A 2AE
[2]Aix-Marseille University, GREQAM-EHESS, 2 rue de la Charité 13002 Marseille
[3]Oxford University, Nuffield College, New Road, Oxford, OX1 3NF

Abstract

Specific functional forms are often used in economic models of distributions; goodness-of-fit measures are used to assess whether a functional form is appropriate in the light of real-world data. Standard approaches use a distance criterion based on the EDF, an aggregation of differences in observed and theoretical cumulative frequencies. However, an economic approach to the problem should involve a measure of the information loss from using a badly-fitting model. This would involve an aggregation of, for example, individual *income* discrepancies between model and data. We provide an axiomatisation of an approach and applications to illustrate its importance.

- Keywords: goodness of fit, discrepancy, income distribution, inequality measurement

- JEL Classification: D63, C10

- Correspondence to: F. A. Cowell, STICERD, LSE, Houghton St, London WC2A 2AE. (f.cowell@lse.ac.uk)

1 Introduction

One of the standard tasks in distributional analysis involves finding a method of judging whether two distributions are in some sense close. The issue arises in the context of the selection of a suitable parametric model and in the context of comparing two empirical distributions. What constitutes a "satisfactory" fit? Obviously one could just apply a basket of standard goodness-of-fit measures to choose among various fits of a given income distribution. But on what criteria are such measures founded and are they appropriate to conventional economic interpretations of distributions? The question is important because choosing the wrong fit will lead to not only to incorrect estimates of key summary measures of the distributions but also to misleading interpretations of distributional comparisons. A variety of measures of goodness of fit have been proposed (Cameron and Windmeijer 1996, 1997, Windmeijer 1995), but the focus in the literature has been on identifying a particular goodness-of-fit measure as a statistic which seems to suit a specific empirical model rather than focusing on their economic interpretation. This paper will examine the problems presented by standard measures of goodness-of-fit for models of distribution and how conventional approaches may give rather misleading guidance. It also suggests an approach to the goodness-of-fit problem that uses standard tools from the economic analysis of income distributions.

As a principal example consider the modelling of empirical income distributions. They are of special interest not just because of their distinctive shape (heavy tailed and right-skewed) but especially because of their use in applied welfare-economic analysis. Income distribution matters for evaluation of economic performance and for policy design because the criteria applied usually take into account inequality and other aspects of social welfare. So a "good" model of the size distribution of income should not only capture the shape of the empirical distribution but also be close to it in a sense that is consistent with the appropriate social-welfare criteria. Obviously the purpose of a goodness-of-fit test is to assess how well a model of a distribution represents a set of observations, but conventional goodness-of-fit measures[1] are not particularly good at picking up the distinctive shape characteristics of income distribution (as we will see later) nor can they be easily adapted to take into account considerations of economic welfare.

In this paper we examine an alternative approach that addresses these questions. The approach is based on standard results in information theory that allow one to construct a distance concept that is appropriate for

[1]There is a huge statistics literature on the subject – see for example d'Agostino and Stephens (1986).

characterising the discrepancies between the empirical distribution function and a proposed model of the income distribution. The connection between information theory and social welfare 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 between information theory and the analysis of income distributions. Section 3 builds on this to introduce the proposed approach to goodness-of-fit. Section 4 sets out a set of principles for distributional comparisons in terms of goodness of fit and show how these characterise a class of measures. Section 5 performs a set of experiments and applications using the new proposed measures and compares them with standard measures in the literature. 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, in section 3, use this approach to discrepancies between distributions in order to formulate an approach to the goodness-of-fit problem.

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[2] where g is any convex function with $g(1) = 0$ (Khinchin 1957). An important special case is given by $g(f) = \frac{1}{\alpha-1}\left[1 - f^{\alpha-1}\right]$ 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.

[2] 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).

It is clear that the function $s(\cdot)$ has the same properties as the regular density function $f(\cdot)$:

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

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

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

and for $\alpha = 1$ is given by (13); note that $J_\alpha(\mathbf{x}, \mathbf{y}) \geq 0$ for arbitrary $\mathbf{x}$ and $\mathbf{y}$.[3] The family (15) represents an aggregate measure of *discrepancy* between two distributions on which we will construct an approach to the goodness-of-fit problem. Again, for a discrete distribution with n point masses, it

[3]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.

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.

Note that if s_2 represents a distribution of perfect equality then (15) becomes the class of generalised-entropy inequality measures: just as the generalised-entropy measures can be considered as the average (signed) distance of an income distribution from perfect equality (Cowell and Kuga 1981), so (15) captures the average distance of an income distribution s_1 from a reference distribution s_2.

3 An approach to goodness-of-fit

The analysis in section 2 provides a natural lead into a discussion of the goodness-of-fit question. Of particular interest is the way in which one transforms from divergences in terms of densities (or probabilities) in the case of information theory to divergences in terms of income shares in the case of income-distribution analysis. This provides the key to our new approach as can be seen from a simple graphical exposition of the goodness-of-fit problem.

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

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

where the x_i are the ordered sample observations and ι is an indicator function such that $\iota(S) = 1$ if statement S is true and $\iota(S) = 0$ otherwise. Figure 1 depicts an attempt to model six data points (on the x axis) with a continuous distribution F_*. The EDF approach computes the differences between the modelled cumulative distribution $F_*(x_i)$ at each data point and the actual cumulative distribution $\hat{F}(x_i)$ and then aggregates the values $\hat{F}(x_i) - F_*(x_i)$; Figure 1 shows one such component difference for $i = 3$. It is difficult to impute economic meaning to such differences and the method of aggregation is essentially arbitrary in economic terms.

However, it is usually the case that economists are more comfortable working within the space of incomes: in the economics literature it is common practice to evaluate whether quantitative models are appropriate by using some loss function defined on real-world data and the corresponding values generated by the model.[4] We adopt a method that is in the same spirit here.

[4] As examples consider (1) the evaluation of performance of general equilibrium models (**?** p153) and (2) the evaluation of the performance of forecasting models (**?**, **?**).

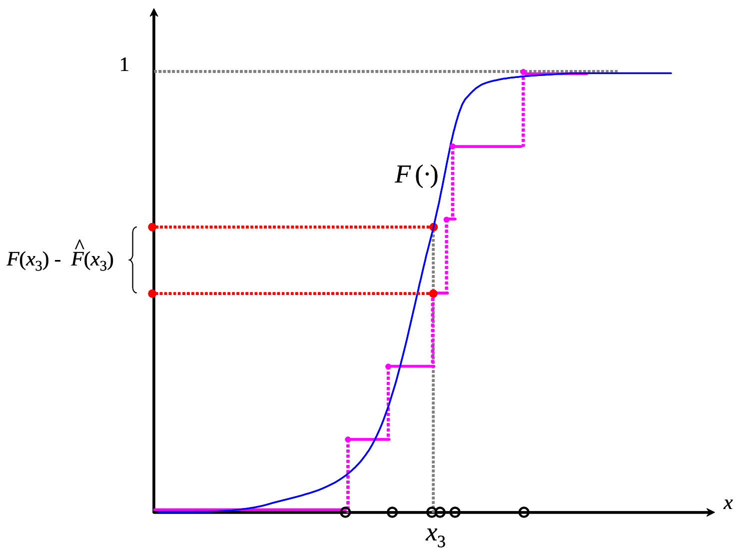

Figure 1: The EDF approach

Instead of the EDF approach we propose using its "dual" where, for each adjusted sample proportion $\frac{i}{n+1}$, $i = 1, .., n$,[5] we compute the corresponding quantile for the theoretical distribution

$$y_i = F_*^{-1}\left(\frac{i}{n+1}\right) \tag{16}$$

and compare the resulting $\mathbf{y}$ vector and the corresponding sample $\mathbf{x}$ vector – see Figure 2 for the case that corresponds to the EDF example in Figure 1. The problem then is similar to that of specifying a loss function in other economic contexts: it amounts to finding a suitable way of comparing the discrete distributions $\mathbf{x}$ and $\mathbf{y}$ in a way that makes economic sense. In section 2 we presented measures of inequality and distributional change obtained by a change of variables from entropy measures. This is equivalent to reasoning in terms of the quantile approach from Figure 2 rather than in terms of the EDF approach from Figure 1. The remaining problem then amounts to finding a suitable way of comparing the discrete distributions $\mathbf{x}$ and $\mathbf{y}$ in a way that makes sense in terms of welfare economics. Suppose the observed distribution is $\mathbf{x}$ and one proposes a model $\mathbf{y}$, where $\mathbf{y} = \mathbf{x} + \Delta\mathbf{x}$. How much does it "matter" that there is a discrepancy $\Delta\mathbf{x}$ between $\mathbf{x}$ and $\mathbf{y}$? The standard

[5] Note that we use $\frac{i}{n+1}$ rather than $\frac{i}{n}$ to avoid an obvious problem where $i = n$. Had we used $\frac{i}{n}$ in (16) then y_n would automatically be set to $\sup(X)$ where X is the support of F_*.

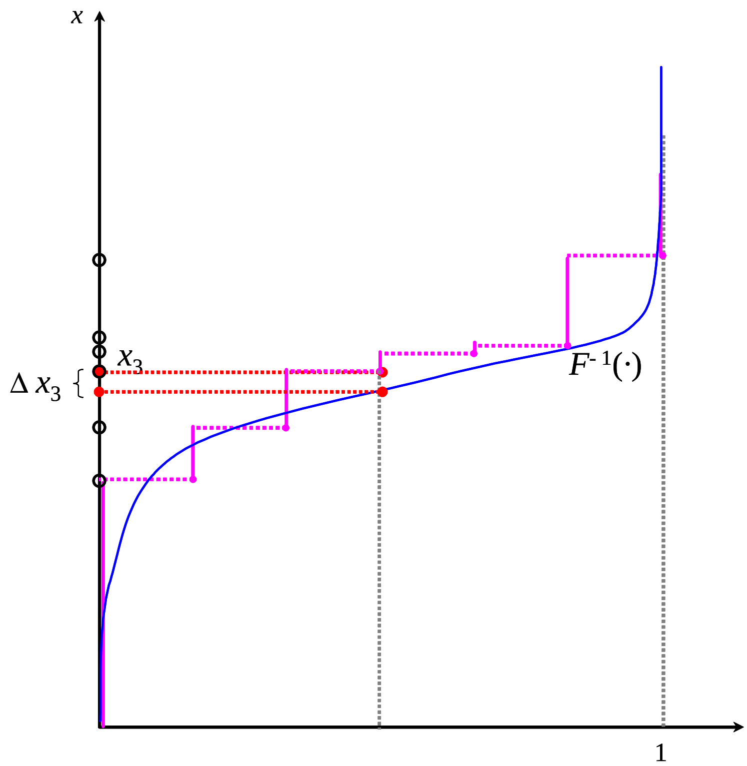

Figure 2: Quantile approach

approach in economics is to look at some indicator of welfare loss. If we were thinking about income distribution in the context of inequality it might also make sense to quantify the discrepancy in terms of inequality change. We may distinguish three separate approaches: welfare loss, inequality change, distributional change. In this section, we consider these three approaches to select an appropriate one and thus propose a new measure of goodness-of-fit based on entropy.

3.1 Welfare loss

Suppose we characterise the social welfare associated with a distribution $\mathbf{x}$ as a function $W : \mathbb{R}^n \to \mathbb{R}$ that is endowed with appropriate properties.[6] If W is differentiable then the change in social welfare in going from distribution

[6] See for, example, Blackorby and Donaldson (1978). The appropriate properties for W would include monotonicity and Schur-concavity.

$\mathbf{x}$ to distribution $\mathbf{y}$ is:

$$W(\mathbf{y}) - W(\mathbf{x}) \simeq \sum_i \frac{\partial W(\mathbf{x})}{\partial x_i} \Delta x_i. \tag{17}$$

Could the welfare difference (17) be used as a criterion of whether $\mathbf{y}$ is "nearer" to $\mathbf{x}$ than some other distribution $\mathbf{y}'$? There are at least two objections. First, it is usually assumed that welfare is ordinal so that W could be replaced by the function $\tilde{W}$ where $\tilde{W} := \varphi(W)$ where φ is an arbitrary monotonic increasing function; if so the expression in (17) is not well-defined as a loss function. Second, the standard assumption of monotonicity means that, for all $\mathbf{x}$, $\frac{\partial W(\mathbf{x})}{\partial x_i} > 0$ so that it is easy to construct an example with $\Delta\mathbf{x} \neq 0$ such that (17) is zero; for instance one could have Δx_i arbitrarily large and positive and Δx_j correspondingly large and negative. But one would hardly argue that $\mathbf{y}$ was a good fit for $\mathbf{x}$. One could sidestep the first objection by using money-metric welfare, in effect taking equally-distributed equivalent income as the appropriate cardinalisation of social welfare, but the second objection remains.

3.2 Inequality change

Suppose instead that we use an inequality index I as a means of characterising an income distribution. Then we have

$$I(\mathbf{y}) - I(\mathbf{x}) \simeq \sum_i \frac{\partial I(\mathbf{x})}{\partial x_i} \Delta x_i. \tag{18}$$

Could the inequality-difference be used as a criterion for judging the "nearness" of $\mathbf{y}$ to $\mathbf{x}$? Essentially the same two objections apply as in the case of welfare change. First, I usually only has ordinal significance so that the measure (18) still depends on cardinalisation of I. Second, consider the standard property of I, the principle of transfers which requires that

$$\frac{\partial I(\mathbf{x})}{\partial x_i} - \frac{\partial I(\mathbf{x})}{\partial x_j}$$

be positive if $x_i > x_j$. Now take also two other incomes $x_h > x_k$ where also $x_k > x_i$. Clearly one can construct $\Delta\mathbf{x} \neq 0$ such that (18) is zero and that the mean of $\mathbf{x}$ remains unchanged. For example let $\Delta x_i = -\Delta x_j = \delta > 0$ and $\Delta x_h = -\Delta x_k = \delta' < 0$: an inequality-increasing income change at the bottom of the distribution (involving i and j) is accompanied by an inequality-decreasing income change further up the distribution (involving h

and k). Evidently δ and δ' may be chosen so that I remains unchanged, and the values of δ and δ' could be large (substantial "blips" in the distribution). Nevertheless the inequality-difference criterion would indicate that $\mathbf{y}$ is a perfect fit for $\mathbf{x}$.

3.3 Distributional change

To see the advantage of this approach let us first re-examine the inequality-difference approach. Consider the effect on (18) of replacing $\mathbf{y}$ by another distribution $\mathbf{y}'$, where

$$\begin{aligned} y'_k &= y_k + \delta \\ y'_j &= y_j - \delta \\ y'_i &= y_i \qquad\qquad \text{if } i \neq j, k \end{aligned}$$

If we were to use the generalised-entropy index (7) then evidently this would be:

$$\Delta_{jk}\left(I_\alpha\left(\mathbf{y}\right) - I_\alpha\left(\mathbf{x}\right)\right) = \frac{1}{n\left[\alpha-1\right]}\left[\left[\frac{y_k}{\mu_1}\right]^{\alpha-1} - \left[\frac{y_j}{\mu_1}\right]^{\alpha-1}\right] \tag{19}$$

where the operator Δ_{jk} is defined by

$$\Delta_{jk}\left(\theta\right) := \frac{d\theta}{dy_k} - \frac{d\theta}{dy_j}.$$

Clearly (19) is positive if and only if

$$\frac{y_k}{y_j} > 1 \tag{20}$$

In other words the change $\mathbf{y} \to \mathbf{y}'$ results in an increase in the inequality-difference as long as y_k is greater than y_j, *irrespective of the value of the vector* $\mathbf{x}$.

Now consider the way the distributional-change measure works when $\mathbf{y}$ is replaced by $\mathbf{y}'$. From (15) we have:

$$\begin{aligned} \Delta_{jk}\left(J_\alpha\left(\mathbf{x},\mathbf{y}\right)\right) &= \frac{1-\alpha}{n\alpha\left[\alpha-1\right]\mu_2}\left[\left[\frac{x_k}{\mu_1}\right]^{\alpha}\left[\frac{y_k}{\mu_2}\right]^{-\alpha} - \left[\frac{x_j}{\mu_1}\right]^{\alpha}\left[\frac{y_j}{\mu_2}\right]^{-\alpha}\right] \\ &= -\frac{1}{n\alpha\mu_1^{\alpha}\mu_2^{1-\alpha}}\left[\left[\frac{x_k}{y_k}\right]^{\alpha} - \left[\frac{x_j}{y_j}\right]^{\alpha}\right] \end{aligned} \tag{21}$$

So $\Delta_{jk}\left(J_\alpha\left(\mathbf{x},\mathbf{y}\right)\right) > 0$ if and only if $\frac{y_k}{x_k} > \frac{y_j}{x_j}$ or equivalently

$$\Delta_{jk}\left(J_\alpha\left(\mathbf{x},\mathbf{y}\right)\right) > 0 \Leftrightarrow \frac{y_k}{y_j} > \frac{x_k}{x_j} \tag{22}$$

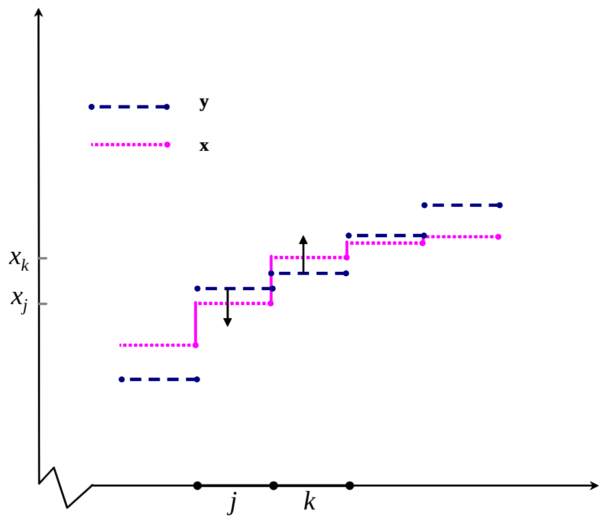

Figure 3: Does the change in $\mathbf{y}$ move one closer to $\mathbf{x}$?

In other words the change $\mathbf{y} \to \mathbf{y}'$ results in an increase in the distributional-change measure as long as the proportional gap between y_k and y_j is greater than the proportional gap between x_k and x_j.

The point is illustrated in Figure 3 which shows part of the quantile representation of the goodness-of-fit approach introduced in Figure 2. Suppose the distribution $\mathbf{y}$ is used as a model of the observed distribution $\mathbf{x}$; for the purposes of the example $I(\mathbf{y}) > I(\mathbf{x})$. For the particular values of j and k chosen it is evident that $x_k > y_k > y_j > x_j$ so that $y_k/y_j < x_k/x_j$. Now consider a perturbation in $\mathbf{y}$ as indicated by the arrows. According to the criterion (22) the distributional-change measure must fall with this perturbation: it appears to accord with a common-sense interpretation of an improvement in goodness-of-fit. But, by construction, the perturbation is a mean-preserving spread of so that inequality $\mathbf{y}$ must increase by the principle of transfers; so according to the inequality-change criterion (18, 20) the fit would have become worse!

It appears that (22) is the appropriate criterion for capturing goodness-of-fit rather than (20) since it incorporates information about the relevant incomes in *both* $\mathbf{x}$ and $\mathbf{y}$ distributions and is independent of information about irrelevant incomes. We will examine this more carefully in section 4.

3.4 A measure of goodness-of-fit based on entropy

We pursue the idea of distributional change as a basis for a loss function by making use of the discrepancy measure J_α introduced in (15).

Given that the population size is normalised to 1, we may define the empirical income-share function $\hat{s} : [0,1] \to [0,1]$ as

$$\hat{s}(q) = \frac{\widehat{F}^{-1}(q)}{\int_0^1 \widehat{F}^{-1}(t)\, dt} = \frac{y}{\hat{\mu}} \tag{23}$$

where $\hat{F}^{-1}(\cdot)$ is the inverse of the empirical distribution function $\hat{F}$ and $\hat{\mu}$ is the mean of this distribution. We may use the concept of relative entropy in Section 2.4 to measure the transformed distribution. Instead of considering a pair of density functions f_1, f_2, we consider a pair of income share functions $\hat{s}$, s. This is a similar consideration as what we have done to make a link between α-class entropy measures and generalised entropy inequality measures. The divergence measure (9) can thus be rewritten

$$H_\alpha(\hat{s}, s) = \frac{1}{\alpha - 1} \int_0^1 \left[\hat{s}(q)^\alpha\, s(q)^{1-\alpha} - 1\right] dq, \quad \alpha > 0$$

where s is given by (3).

For the goodness-of-fit problem we apply the corresponding discrepancy measure J_α to the case where we have an empirical distribution and a theoretical distribution. Take a sample of size n: for the empirical distribution the shares are given by (23) and for each q the corresponding share in the theoretical distribution F_* is given by

$$s(q) = \frac{F_*^{-1}(q)}{\int_0^1 F_*^{-1}(t)\, dt} = \frac{F_*^{-1}\left(\frac{i}{n+1}\right)}{\mu(F_*)}$$

where $q = \frac{i}{n+1}$ and so we have, as a possible goodness-of-fit criterion:

$$J_\alpha = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left[\frac{x_{(i)}}{\hat{\mu}}\right]^\alpha \left[\frac{F_*^{-1}\left(\frac{i}{n+1}\right)}{\mu(F_*)}\right]^{1-\alpha} - 1\right], \ \alpha \neq 1 \tag{24}$$

$$J_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{x_{(i)}}{\hat{\mu}} \log\left(\frac{x_{(i)}}{\hat{\mu}} \Big/ \frac{F_*^{-1}\left(\frac{i}{n+1}\right)}{\mu(F_*)}\right). \tag{25}$$

where $x_{(1)}, x_{(2)}, ...$ denote the members of the sample in increasing order.

However, this class of goodness-of-fit measures is based on an intuitive comparison with the problem of quantifying distributional change. In fact the goodness-of-fit problem is not exactly the same as distributional change so that it would be inappropriate just to "borrow" the analysis. Accordingly in the next section we examine the fundamentals of the approach.

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

As with the distributional change problem, the goodness-of-fit 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 goodness-of-fit 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[7] $\succeq$ 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 good a fit 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.[8]

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.

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

[8] 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 "$[\mathbf{z} \succeq \mathbf{z}'] \,\&\, [\mathbf{z}' \succeq \mathbf{z}]$".

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 $d(x_i, y_i) < d(x_i', y_i') \Longleftrightarrow \mathbf{z} \succ \mathbf{z}'$.*

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 $y_1 \leq y_2 \leq ... \leq y_n$.[9] 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 goodness-of-fit and the fit 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 (Perfect local fit) *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 a perfect fit 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 goodness-of-fit.

Theorem 6 *Given Axioms 1 to 5 (a) $\succeq$ is representable by the continuous function given by*

$$\sum_{i=1}^{n} \phi_i(z_i), \forall \mathbf{z} \in Z^n \tag{26}$$

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)

$$\phi_i(x, x) = a_i + b_i x \tag{27}$$

[9] 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.

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 (26) and it is clear that $\mathbf{z} \sim \mathbf{z}'$ if and only if

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

which can only be true if

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

for arbitrary x_i and δ. This is a standard Pexider equation and its solution implies (27). ■

Corollary 7 *Since $\succeq$ is an ordering it is also representable by*

$$\phi\left(\sum_{i=1}^{n} \phi_i\left(z_i\right)\right) \tag{28}$$

where, ϕ_i is defined as in (26), (27). and $\phi : \mathbb{R} \to \mathbb{R}$ continuous and strictly monotonic increasing.

This additive structure means that we can proceed to evaluate the goodness-of-fit problem 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 9 shows it is enough to induce a rather specific structure on the function representing $\succeq$.

Axiom 8 (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 9 *Given Axioms 1 to 8 $\succeq$ is representable by*

$$\phi\left(\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right)\right) \tag{29}$$

where h_i is a real-valued function.

Proof. Using the function Φ introduced in the proof of Theorem 6 Axiom 8 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 (26), we have for all :

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

or, equivalently

$$\phi_i(tx_i, ty_i) = \psi(t)\,\phi_i(x_i, y_i)\,, i = 1, ..., n. \tag{30}$$

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

$$\phi_i(x_i, y_i) = x_i^c h_i\left(\frac{x_i}{y_i}\right). \tag{31}$$

From (27) and (31) it is clear that

$$\phi_i(x_i, x_i) = x_i^c h_i(1) = a_i + b_i x_i, \tag{32}$$

which implies $c = 1$. Putting (31) with $c = 1$ into (28) 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 and goodness-of-fit

We now focus on the way in which one compares the (x, y) discrepancies in different parts of the income distribution. The form of (29) 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 10 (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 distributional fits $\mathbf{z}_0$ 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 distributional fits $\mathbf{z}$ and $\mathbf{z}'$ will also be equivalent.[10]

Theorem 11 *Given Axioms 1 to 10* $\succeq$ *is representable by*

$$\phi\left(\sum_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha}\right) \tag{33}$$

where $\alpha \neq 1$ *is a constant.*[11]

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

$$r = \psi\left(\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right)\right) \tag{34}$$

where ψ is the solution in r to

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

In (35) can take the x_i as fixed weights. Using Axiom 10 in (34) requires

$$tr = \psi\left(\sum_{i=1}^{n} x_i h_i\left(t\frac{x_i}{y_i}\right)\right), \text{ for all } t > 0. \tag{36}$$

Using (35) we have

$$\sum_{i=1}^{n} x_i h_i\left(t\psi\left(\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right)\right)\right) = \sum_{i=1}^{n} x_i h_i\left(t\frac{x_i}{y_i}\right) \tag{37}$$

[10] Also note that Axiom 10 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.

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

Introduce the following change of variables

$$u_i := x_i h_i \left(\frac{x_i}{y_i} \right), i = 1, ..., n \tag{38}$$

and write the inverse of this relationship as

$$\frac{x_i}{y_i} = \psi_i \left(u_i \right), i = 1, ..., n \tag{39}$$

Substituting (38) and (39) into (37) we get

$$\sum_{i=1}^{n} x_i h_i \left(t\psi \left(\sum_{i=1}^{n} u_i \right) \right) = \sum_{i=1}^{n} x_i h_i \left(t\psi_i \left(u_i \right) \right). \tag{40}$$

Also define the following functions

$$\theta_0 \left(u, t \right) := \sum_{i=1}^{n} x_i h_i \left(t\psi \left(u \right) \right) \tag{41}$$

$$\theta_i \left(u, t \right) := x_i h_i \left(t\psi_i \left(u \right) \right), i = 1, ..., n. \tag{42}$$

Substituting (41),(42) into (40) 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

$$h_i \left(t \frac{x_i}{y_i} \right) = \frac{b_i \left(t \right)}{x_i} + B \left(t \right) h_i \left(\frac{x_i}{y_i} \right), i = 1, ..., n \tag{43}$$

From Eichhorn (1978), Theorem 2.7.3 the solution to (43) is of the form

$$h_i \left(v \right) = \begin{array}{ll} \beta_i v^{\alpha - 1} + \gamma_i, & \alpha \neq 1 \\ \beta_i \log v + \gamma_i & \alpha = 1 \end{array} \tag{44}$$

where $\beta_i > 0$ is an arbitrary positive number. Substituting for $h_i(\cdot)$ from (44) into (9) 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 (33), $\sum_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha}$, 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

$$J_{\alpha}(\mathbf{x}, \mathbf{y}) := \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left[\frac{x_i}{\mu_1} \right]^{\alpha} \left[\frac{y_i}{\mu_2} \right]^{1-\alpha} - 1 \right], \tag{45}$$

This normalised version of the goodness-of-fit index can be implemented straightforwardly for a proposed model of income distribution. Of course this would require the choice of a specific value or values for the parameter α in (45) according to the judgment that one wants to make about the relative importance of discrepancies in different parts of the income distribution: choosing a large positive value for α would put a lot of weight on discrepancies in the upper tail; choosing a substantial negative value would put a lot of weight on lower-tail discrepancies – see the discussion on page 4.

5 Implementation

We now look at the practicalities of the class of measures J_{α}, interpreted as discrepancy measures (section 5.1) and as goodness-of-fit measures (section 5.2)..

5.1 J as a measure of discrepancy

In empirical studies, χ^2 and EDF are commonly used as goodness-of-fit measures when the income distribution is estimated from a parametric function; summary statistics such as inequality measures are then computed from this estimated income distribution. As we saw in section 3.2, goodness-of-fit and inequality measures are not based on similar foundations and can thus lead to contradictory results. By contrast, J measures and generalized entropy inequality measures have similar foundations and should provide consistent results. We show this in this section with an experiment using the J index as a measure of discrepancy.

Take three income distributions, constructed such that f_1 and f_0 are similar in high incomes, while f_2 and f_0 are similar in low incomes. These density functions, defined as mixtures of three Lognormal distributions, are

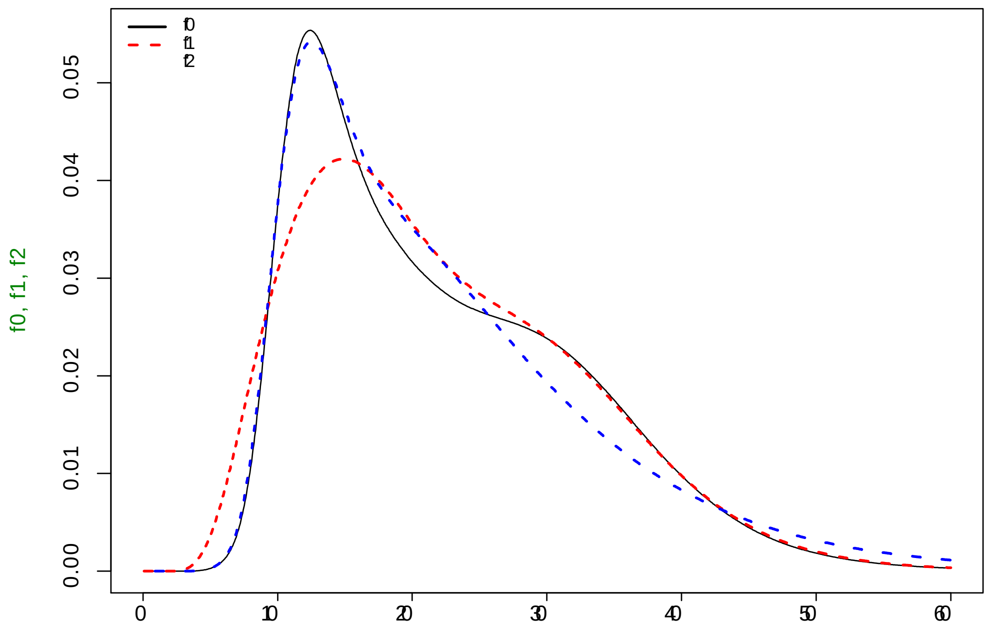

Figure 4: The three lognormal mixtures f_0, f_1, f_2

plotted in figure 4.[12]In this experiment, we address the question: *which of* f_1 *and* f_2 *shows the smaller divergence from* f_0*?*

A standard approach to this question is to choose a measure of Goodness-of-Fit and to minimize it. We compute χ^2 and ω^2, the Cramér-von Mises (EDF) measures,[13] results are given in the right-hand side of Table 1. Minimizing these two measures, we conclude that the discrepancy between f_2 and f_0 is smaller than between f_1 and f_0. What if, instead, we used inequality as a measure of discrepancy between distributions? Table 2 reports inequality measures (7), $\alpha = 0, 1$ for the three distributions. For both values of α we get the opposite of what we concluded from χ^2 and ω^2: in inequality terms distribution f_1 is "closer" to f_0 than f_2.

[12]We have $f_k(x) = p_1\, \lambda(x; \mu_1, \sigma^2_{1k}) + p_2\, \lambda(x; \mu_2, \sigma^2_{2k}) + p_3\, \lambda(x; \mu_3, \sigma^2_{3k})$, where λ represents the lognormal density function, $p_1 = p_3 = 0.2$, $p_2 = 0.6$, $\mu_1 = 2.5$, $\mu_2 = 3$, $\mu_3 = 3.5$ and $\sigma_2 = 0.4$. The differences between the three distributions come only from σ^2_{1k} and σ^2_{3k}: we have chosen $f_0(x) : \sigma_{10} = 0.2, \sigma_{30} = 0.2$; $f_1(x) : \sigma_{11} = 0.4, \sigma_{31} = 0.21$; and $f_2(x) : \sigma_{12} = 0.21, \sigma_{32} = 0.35$.

[13]For a comprehensive treatment of standard EDF criteria see Anderson and Darling (1954), **?**).

	$J_\alpha\times\ 10^2$				
α	f_1	f_2		f_1	f_2
-1.0	0.079	0.191	χ^2	0.058679	0.048541
-0.5	0.076	0.195	ω^2	3.556511	2.421263
0.0	0.0742	0.1989			
0.5	0.0720	0.2028			
1.0	0.0699	0.2070			

Note: computations for 10,000 simulated data points in f_0, f_1 and f_2

Table 1: Comparing f_1 and f_2 as approximations to f_0: J, χ^2 and ω^2 statistics

	f_0	f_1	f_2
I_0	0.104396	0.113890	0.120640
I_1	0.101353	0.106494	0.121767

Table 2: Comparing f_1 and f_2 as approximations to f_0: Inequality measures

Of course, using the difference between two inequality indexes as a measure of discrepancy is inappropriate, as we saw in section 3.2. The left-hand side of Table 1 presents values of the appropriate discrepancy measures J (15), for various values of α. Clearly the discrepancy with f_0 is always larger in the case of f_2 than f_1 – the opposite conclusion of what one obtains with χ^2 and ω^2, but in accordance with inequality measurement.

What is also interesting to note is how the extent of the discrepancies vary between the estimates of J with the different values of α.We find that the higher the value of α, the closer the approximation of f_1 to f_0 and the worse is that of f_2. With α representing the sensitivity parameter of the inequality index involved, (in other words, with a higher value of α giving greater weight to higher incomes), this allows for two separate interpretations. On the one hand, one may read this result as suggesting that for income distribution estimations with the purpose of focusing on incomes of the poor, the choice of a low value of α is sensible. On the other hand, if one is interested in the distribution of wealth or incomes in the upper tail, higher values of α are particularly relevant.

5.2 J as a goodness-of-fit measure

Let us compare performance of the statistic J_α with that of conventional goodness-of-fit criteria when applied to UK income data.[14] The empirical

[14]The application uses the "before housing costs" income concept of Households Below Average Incomes 2004-5 (?), deflated and equivalised using the OECD equivalance scale,

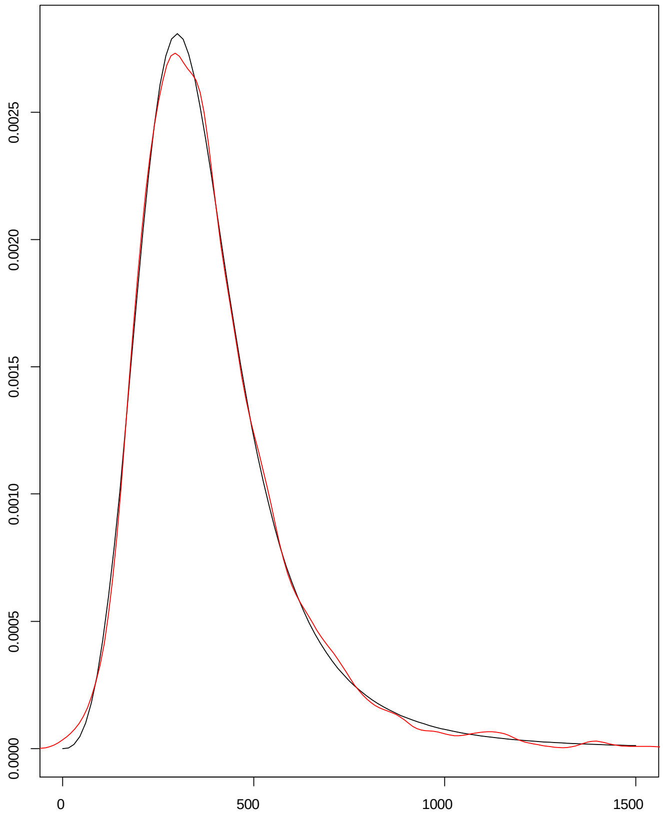

Figure 5: Kernel density of the observed distribution (red) and the Singh Maddala fit (black)

distribution $\hat{F}$ consists of 3858 observations and has mean and standard deviation $(398.28, 253.75)$. Consider as a model for these data the Singh-Maddala distribution

$$F_{\mathrm{SM}}(y; a, b, c) = 1 - \frac{1}{[1 + ay^b]^c}. \tag{46}$$

The parameters are unknown so we use the maximum-likelihood estimates $\left(\hat{a}, \hat{b}, \hat{c}\right) = (5.75554E^{-10},\ 3.6303,\ 1.0106)$; an impression of the suitability of the SM model with these parameters is given by Figure 5.[15] Formally, we test

for the cohort of ages 21-45, couples with and without chidren, excluding households with self-employed individuals. The variable used in the dataset is oe_bhc. We exclude households with self-employed individuals as reported incomes are known to be misrepresented.

[15] We also tested the Dagum and Lognormal distributions; the Singh Maddala distribution appears as the best fit for these data.

α	$J_\alpha \times 10^2$	p (%)	α	$J_\alpha \times 10^2$	p (%)			p (%)
-5	2.0313	0.00	0.2	0.1288	5.41	χ^2	54.4417	32.33
-2	0.1480	1.90	0.5	0.1312	6.01	ω^2	0.04766	29.43
-1	0.1276	3.80	0.7	0.1332	7.21			
-0.7	0.1263	4.00	1	0.1366	6.71			
-0.5	0.1261	5.41	2	0.1519	8.31			
-0.2	0.1267	5.11	5	0.2394	10.01			
0	0.1276	5.31						

Table 3: The SM model for UK income distribution: J, χ^2 and ω^2 statistics

the hypothesis $\mathrm{H}_0 : J_\alpha(\mathbf{x}, \mathbf{y}) = 0$ against the alternative $\mathrm{H}_1 : J_\alpha(\mathbf{x}, \mathbf{y}) \neq 0$ where $\mathbf{x}$ is the sample vector of incomes and $y_i = F_{\mathrm{SM}}^{-1}\left(\frac{i}{n+1}; \hat{a}, \hat{b}, \hat{c}\right)$. Table 3 reports the point estimates of J_α along with the associated probability p of a type-1 error estimated using bootstrap methods; the right-hand columns present the corresponding estimates and p-values of χ^2 and ω^2.[16]

While χ^2 and ω^2 indicate that the Singh-Maddala distribution is satisfactory (high p-values), the J_α criterion reveals a richer story. Observe that the p-values rise with α: in other words it is appropriate to accept F_{SM} as a suitable fit to $\hat{F}$ if one uses a criterion that assigns higher weight to discrepancies between model and data at high incomes. However, H_0 should be rejected for $\alpha < 1$ so, for a "bottom-sensitive" goodness-of-fit criterion, F_{SM} would be regarded as unsatisfactory. This conclusion regarding the role of α is similar to that noted at the end of Section 5.1 when considering the performance of the J index as a discrepancy measure for the experiment involving mixtures of lognormal distributions.

6 Conclusion

Why do economists want to use goodness-of-fit criteria? The principal application of such criteria is surely in evaluating the empirical suitability of a statistical model used in an economic context – perhaps the outcome of income or wealth simulations or the characterisation of an equilibrium distribution of an economic process. It seems reasonable to use a fit criterion that is in some way based on economic principles rather than just relying one or

[16] For a given α a bootstrap sample is drawn from $F_{\mathrm{SM}}(\cdot; \hat{a}, \hat{b}, \hat{c})$ and from this an estimate $\tilde{J}_\alpha$ is obtained. We perform 999 such bootstrap samplings with replacement to obtain 999 estimates of $\tilde{J}_\alpha$. The p-value is then estimated as the proportion of $\tilde{J}_\alpha s$ that exceed J_α in value.

two off-the-shelf statistical tools.

Our approach – the "dual" to the statistical EDF method – 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 or as goodness-of-fit criteria. 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 fit criteria 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 the question: in which part of the distribution do you want the goodness-of-fit criterion to be particularly sensitive?

Our simulation exercise (in Section 5.1) shows that off-the-shelf tools can be misleading in evaluating discrepancy between distributions but that the J_α indices provide answers that accord with common sense. The application to modelling real data (in Section 5.2) shows that the sensitivity parameter α is crucial to understanding whether the proposed functional form is appropriate. The choice of a fit criterion really matters.

References

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

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

Anderson, T. and D. A. Darling (1954). A test of goodness of fit. *Journal of the American Statistical Association 49*, 765–759.

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

Blackorby, C. and D. Donaldson (1978). Measures of relative equality and their meaning in terms of social welfare. *Journal of Economic Theory 18*, 59–80.

Cameron, C. and F. Windmeijer (1996). R-squared measures for count data regression models with applications tohealth care utilization. *Journal of Business and Economic Statistics 14*, 209–220.

Cameron, C. and F. Windmeijer (1997). An R-squared measure of goodness of fit for some common nonlinear regressionmodels. *Journal of Econometrics 77*, 329–342.

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

Cowell, F. A. and K. Kuga (1981). Inequality measurement: an axiomatic approach. *European Economic Review 15*, 287–305.

d'Agostino, R. B. and M. A. Stephens (1986). *Goodness-of-fit techniques.* New York: Marcel Dekker.

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.

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

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.

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.

Windmeijer, F. (1995). Goodness-of-fit measures in binary choice models. *Econometric Reviews 14*, 101–116.