Authors: Sanghamitra Bandyopadhyay, Frank Cowell and Emmanuel Flachaire

Affiliation: London School of Economics and University of Paris-1

Short Abstract:

This paper uses inequality-measurement techniques to assess goodness of fit in income-distribution models. It exposes the shortcomings of the use of conventional goodness of fit criteria in the estimation of income and wealth distributions, and proposes a new set of criteria, and new metrics, based on generalised entropy measures. A number of experiments reveal that conventional fit criteria, as in the case of the Kolmogrov-Smirnov, Anderson-Darling or the Cramer von Mises statistics result in worse fit than metrics based on some basic (generalised entropy) measures of inequality. This is revealed to be because of conventional (distance based) goodness-of-fit measures being based on the Euclidean concept of distance, thus giving equal weight to parts of the distribution in which one is not so interested. We then propose a set of conditional entropy metrics, (as opposed to distance measures), as an alternative set of measures. The proposed metrics are compared against a set of distributional-change axioms already established in the inequality literature (Cowell, Review of Econ Studies 1985).

Extended Abstract:

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 welfare-theoretic interpretations of income distributions? The question is important because choosing the wrong fit will lead to not only to incorrect estimates of the extent of inequality 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 measure of goodness-of-fit as a statistic which seems to suit a specific empirical model rather than focusing on the conventional welfare-theoretic interpretations of income distributions. This paper will examine the problems that one has in using the standard measures goodness of fit for estimating income distribution models in particular and how conventional approaches may give rather misleading guidance. We propose a set of goodness of fit measures using generalised entropy measures which take into account the particular nature of income distributions models.

Estimation of income distributions lend themselves as interesting cases because of the particular shape they take. Typically bottom heavy with long tapering tails on the right, they are highly skewed in their distribution.

Standard measures of goodness of fit, however, are not constructed to empirically recognise particular shapes of the densities whose shapes they are to test for. Choosing a certain fit of the income distribution using any of the standard measures, like the Cramér-von Mises or Kolmogrov-Smirnov or Anderson-Darling statistics, will give us an erroneous fit.

This is because these measures are structured to give the same weight to all observations in the distribution. The goodness of fit measures which are typically used in the statistical literature (i.e. the Cramér-von Mises or Kolmogrov-Smirnov or Anderson-Darling statistics, give equal weight to parts of the distribution which do not require so (like the tapering, sparsely populated upper tail, and vice versa). What we show in this paper is that this can lead to contradicting results to those derived using a simple generalised entropy measures. Entropy based measures are designed to take di¤erential weighting into account. The results of the experiment performed in the paper using a generalised entropy measures as a goodness of fit statistic reveal that

the fit significantly improves. This leaves us with the question of whether Euclidean distance based measures, such as the Kolmogrov-Smirnov, or the Cramér-von Mises measure are the correct metric to be used.

Entropy measures, closely related to such Euclidean distance-based measures improve over these metrics, as they convey more information about the discrepancy that is being measured. Euclidean (or Hilbert) and Kolmogrov distances are simply n-dimensional aggregated discrepancies for the mapping of f1 to f2. Entropy measures, on the other hand, conceptualise the discrepancy as one associated with the lack of information associated with the event yk. On employing this probabilistic approach, one can weight (and thus obtain an expected measure of information) the discrepancies with their probabilities of occurrences. It provides itself as a metric which is particularly intended to measure discrepancies. This is pertinent for the fit of income distributions due to their particularly skewed shapes.

In this paper we propose some goodness of fit measures using some information theoretic generalised entropy measures for income distribution estimations. The use of entropy measures in econometrics has been relatively new, but has been increasing in its applications widely. Using entropy measures for goodness of fit of regression models has been however limited, for the exception of some generalised work of obtaining an equivalent of the R squared.(Cameron and Windmeijer 1997). We will first perform an experiment to highlight the extent to which our estimates of inequality across the same income distribution are significantly different for different empirical distributions estimated using standard measures of goodness of fit. Our results show that using the Theil index of inequality as a metric of goodness of fit, the empirically fitted income distribution comes closest to the true population income distribution.

We then propose a number of entropy based distance measures, which satisfy all the necessary properties of a distance metric, and derive a measure of distributional change based on (Cowell 1980) and (Cowell 1985). An axiomatic framework based on Cowell (1985) is derived on the basis of which we can establish the proposed measure as a satisfactory distance measure.

The axiomatisation follows a characterisation of the concept of distributional change and incorporates standard structural assumptions such as scale invariance and decomposability: it has much in common with the concepts underlying conditional entropy.