THE ECONOMIC JOURNAL

The Economic Journal, **129** (*July*), 2216–2239 DOI: 10.1111/ecoj.12637
Advance Access Publication Date: 8 January 2019

INEQUALITY INDICES AS TESTS OF FAIRNESS*

Ravi Kanbur and Andy Snell

Inequality indices are traditionally interpreted as measures of deviations from equality. This article interprets them instead as statistical tests for a null of fairness within well-defined income generating processes. We find that the likelihood ratio (LR) test for fairness versus unfairness within two such processes are proportional to Theil's first and second inequality indices respectively. The LR values may be used either as a test statistic or to approximate a Bayes factor that measures the posterior probabilities of the fair version of the processes over that of the unfair. We also apply this perspective to measurement of inequality of opportunity.

The standard procedure for measuring income inequality in a society is to take its snapshot observed distribution of income and to calculate an inequality index from it.[1] Such indices have also been interpreted as a measure of welfare loss entailed in departures from equality of outcomes, for egalitarian social welfare functions defined on the distribution of outcomes.[2] However, this procedure faces the well-known critique that the observed distribution is nothing but a snapshot outcome of a process, and that it is the process which matters for normative assessment. In particular, it is the fairness of the underlying process which is held to be the appropriate normative standard, not whether the observed inequality of outcomes is high or low.[3] But if we take the process versus outcomes critique seriously, does this mean that we stop calculating total inequality as something which has normative validity in and of itself?

We argue in this article that in the presence of luck as a determinant of the income distribution, overall indices of inequality still maintain their relevance, but now as statistical tests of fairness, in the sense of whether individuals draw from the same or very different lotteries in life. In particular, we motivate the use of Theil's two canonical indices of inequality, including the increasingly well-used mean log deviation (MLD) measure, as tests of a null of a 'fair' income process versus an alternative of an unfair one in the same class.

Let us start with a simple motivation for the core arguments of this article. Suppose that there are two individuals and a pot of money. Suppose that they agree to toss for the pot with a coin that is fair and known to be fair. At the end of the toss, one individual has the whole pot; the other nothing. The income distribution will be very unequal ex post, but ex ante there is perfect equality because each individual had an equal chance of winning—they faced the same lottery of life. It could be said that the process was fair even though the outcomes were unequal. It could further be argued that it is highly misleading to calculate standard inequality indices from the observed distribution of income in this situation because what is important normatively is the fairness of the process, not equality of outcome. Such an argument was made forcefully by Milton Friedman in *Capitalism and Freedom*:

* Corresponding author: Andy Snell, School of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh EH8 9JT, UK. Email: a.j.snell@ed.ac.uk

This paper was received on 17 March 2017 and accepted on 12 June 2018. The Editor was Martin Cripps.

[1] For an overview and survey of standard methods and interpretation, see Cowell (2011).
[2] The classic reference here is Atkinson (1970).
[3] Early critiques are by Sen (1980), Dworkin (1981) and Roemer (1998). A recent survey is by Roemer and Trannoy (2015).

'Another kind of inequality arising through the operation of the market is also required, in a somewhat more subtle sense, to produce equality of treatment. It can be illustrated most simply by a lottery. Consider a group of individuals who initially have equal endowments and who all agree voluntarily to enter a lottery with very unequal prizes. The resultant inequality is surely required to permit the individuals in question to make most of their initial equality.' (Friedman, 1962, p. 162)

The key point is that even though individuals face an identical lottery and thus equal prospects, the outcomes are bound to be unequal as the lottery pays out its unequal returns. With a large number of individuals, the distribution of income will reflect the distribution of lottery outcomes. But this observed distribution of income, and in particular any measure of inequality calculated from it, has no normative significance.

The lottery being identical across individuals is central to the intuition that the observed inequality of outcome is misleading as a guide to fairness of process. Suppose now, however, that there is the possibility that the lotteries individuals face are not identical. Then the underlying process is not fair in this sense—some have a better luck of the draw than others. But how would we know whether we were in the Friedman case of identical lotteries or the case with non-identical lotteries? If we could directly observe the lottery characteristics and choices, that would be the end of the matter. But typically, all we can observe are in fact the outcomes. The task is then to try to infer from these outcomes the nature of the process that generated them. With a large number of individuals we will observe an income distribution at the end of the lottery draw, but this will now be a mixture of distributions reflecting the different lotteries. The question is: were the lotteries identical?

It should be obvious that in order for this question to be meaningful we will have to provide further structure to the process which generates the income distribution. Thus, trivially, we will not be able to reject the hypothesis that all individuals faced the empirically observed income distribution as the identical lottery of life! Section 1 presents two canonical processes within which we can identify fairness, 'lotteries being the same for all individuals', in well specified fashion. We then show that the LR test statistics for fairness so specified are closely linked to two canonical measures of inequality used in the literature, both due to Theil.[4] Of particular interest is that the MLD (Theil's second measure), which is fast becoming the workhorse measure in inequality and inequality of opportunity analysis, can be interpreted as a test statistic for fairness for a class of income generation processes.

It is remarkable that Theil's inequality indices can be reinterpreted as statistical tests of whether the lotteries faced by individuals are identical, of course within a well-specified class of lotteries. But do non-identical lotteries necessarily signify unfairness? Not if the variation in lotteries across individuals is itself justified according to other normative criteria. Roemer (1998) famously crystallised earlier discussion by attributing variation of outcomes to (*i*) circumstance, those factors which are outside the control of the individual and (*ii*) effort, those factors which the individual controls. Thus race, gender, age, parental wealth etc are clearly outside the control of an individual and variations in outcomes attributable to such factors are illegitimate according to Roemer—they constitute 'inequality of opportunity'.[5] Section 2 of this

[4] The original reference is Theil (1967). Theil drew on Shannon (1948) and information theory to develop his measures, and these were in turn drawn from entropy theory. The Theil family of measures is often known as the 'entropy family' of inequality measures. For a modern account, see Cowell (2011).

[5] The 'post-Roemer' literature is huge, including, for example, Bourgignon *et al.* (2006), Corak (2013), Chetty *et al.* (2014) and Kanbur and Stiglitz (2016).

article builds on the results of Section 1 and develops a test for whether lotteries are non-identical across circumstance groups—in other words, a test for inequality of opportunity in lotteries. Roemer (1998) himself develops a metric for measuring inequality of opportunity based on the mean MLD, which has now been applied the world over.[6] But his development is not in the framework of lotteries. Section 3 also relates our results to Roemer's metric for inequality of opportunity.

The development in Sections 1 and 2 is in a classical statistical framework. Section 3 takes a Bayesian perspective and shows that if the null and alternative are treated as equally likely prior 'models' then the Theil indices are also proportional to the log of the relative posterior probabilities of unfairness to fairness under these two processes. Section 4 presents an empirical illustration of the statistical tests developed in Sections 1, 2 and 3, by applying them to data from the National Longitudinal Survey of Youth from the USA.

In the post-Roemer (1998) literature, there are two approaches to developing a metric for inequality of opportunity.[7] One approach is to take the variation in outcome attributable to circumstance variables as being a lower bound estimate of inequality of opportunity. The second approach is to take the variation attributable to effort as an upper bound estimate of inequality of opportunity. The difficulty with the second approach is to find suitable measures of effort in data. A similar issue arises in our lottery framework. Variation in lotteries attributable to effort would be considered legitimate in the framework, but how are we to allow for this variation in our testing? Section 5 presents a route to thinking about this question and sets out conditions under which the form of tests developed in Sections 1, 2 and 3 are preserved, although the critical values to be used in testing for fairness are now different.

Section 6 returns to the canonical lotteries, which are the work horses of the statistical framework in this article. We think of them as being reduced form devices for capturing income-generating processes. The section discusses alternative structural forms that could be consistent with the lotteries framework put forward in Section 1. Section 7 concludes with a broader discussion of fairness and testing for fairness.

1. Two Lotteries and Likelihood Ratio Tests for Fairness

Friedman's example, it will be recalled, is one in which a single lottery is faced by all individuals but, because of the luck of the draw, there is inequality in the observed distribution. Thus all individuals facing the same lottery is a sharp characterisation of 'fairness'. In this section we develop LR tests for the sameness of lotteries across individuals. But it is clear that we will have to do this within a well-specified framework of allowable lotteries. Thus, as already noted, we obviously cannot reject the hypothesis that all individuals faced the empirically observed income distribution as a lottery. In effect, we derive LR test statistics of a null that each individual in an economy draws incomes from the same pdf (same lottery) versus an alternative that they draw from different pdfs. The LR testing procedure requires that all parameters are identified under the null and the alternative and this disciplines the set of alternatives that we may explicitly consider. As a consequence, we confine our attention to single parameter distributions and test the null that

[6] See, for example, Barros *et al.* (2009) and Ferreira and Gignoux (2011).
[7] See Niehues and Peichl (2014)

the relevant parameter is the same for all n individuals in the sample versus the alternative that each individual has a different parameter.[8]

We introduce two income distribution processes that will frame our discussion and analysis of testing for fairness. We envisage them both in terms of 'helicopter drops', but what is being dropped differs between the two cases. In the first case the helicopter hovering over the population drops dollars that attach to individuals at random. In the second case there is initially a uniform flow of dollars to individuals on the ground before the helicopter comes on to the scene. What the helicopter does is to drop *stops* to this flow, which attach to individuals at random.

We refer to the two processes above as the 'helicopter money drop' and the 'helicopter money stop' respectively. These two processes generate lotteries for individuals and we can identify 'fairness' within each process in a specified way. The processes that we introduce are clearly reduced form in nature, although we believe that they also have intuitive appeal in terms of structure. A discussion of economic mechanisms that may generate these processes is taken up in Section 6. We now develop tests for fairness in each case and relate the tests to conventional inequality indices.

1.1. *The 'Helicopter Money Drop'*

We envisage a helicopter dropping Y units of currency (dollars, say, and henceforth referred to as just 'income units') onto a population. The null is that the probability of each person receiving an income unit is equal and the alternative is that these probabilities are unequal.

Let y_i ($i = 1, 2, ..., n$) denote individual i's income, $Y = \sum_{i=1}^{n} y_i$ denote the total number of income units, $s_i = \frac{y_i}{Y}$ the income share of the ith individual and $\mu = \frac{Y}{n}$ be average income per head again measured in income units. Consider the process where each of the Y income units is allocated across the n individuals. Let the probability that individual i receives one unit be p_i. Each unit is assumed to be distributed independently so that the pdf of $y_1, y_2, ..., y_n$ is multinomial with likelihood

$$L(y_1, y_2, ..., y_{n-1}; p_1, p_2, ..., p_n) = \frac{Y!}{y_1!y_2!...y_n!} p_1^{y_1} p_2^{y_2} ... p_n^{y_n},$$

where $p_n = 1 - \sum_{i=1}^{n-1} p_i$ and $y_n = Y - \sum_{i=1}^{n-1} y_i$.

As noted, we consider the null hypothesis that, ex ante, each individual has an equal chance of receiving an income unit. Hence, under H_0 we have the n-1 restrictions

$$H_0: \qquad p_1 = p_2 = ... = p_n \tag{1}$$

and an alternative H_1 that one or more of these restrictions are violated. This null hypothesis encapsulates precisely and in analytical terms what we mean by fairness/equity in the context of the current article. We now show that the LR test is proportional to Theil's first inequality index.

The log likelihood (l) is

$$\ln L(.) = l = \ln(Y!) - \sum_{i=1}^{n} \ln y_i! + \sum_{i=1}^{n} y_i \ln p_i. \tag{2}$$

[8] Under the alternative we assume that each individual draws from the same type of pdf, but with a different parameter. Were we to analyse pdfs that are determined by two or more parameters, the log likelihood would be undefined under the alternative. This composite alternative is restrictive. However, although the test is designed with a specific composite alternative in mind, this does not imply that it will have zero power against more general deviations from the null.

A form for the LR test of (1) is found by comparing the values of (2) obtained when $p_i = \frac{1}{n}$ (the null 'estimates') with that obtained under $p_i = \frac{y_i}{Y}$ (the alternative estimates). Hence we have

$$\begin{aligned} LR_1 &= 2(l_n - l_1) = 2\left(\sum_{i=1}^{n} y_i \ln \frac{y_i}{Y} - \sum_{i=1}^{n} y_i \ln \frac{1}{n}\right) \\ &= 2\left(\sum_{i=1}^{n} y_i(\ln s_i + \ln n)\right) = 2Y\left(\sum_{i=1}^{n} s_i \ln ns_i\right) \\ &= 2Y.T_1 = 2n\mu T_1, \end{aligned}$$

where T_1 is Theil's 'first' index of inequality, $l_n(l_1)$ are the maximised log likelihoods assuming there are n (1) probability parameters and LR_1 is the LR test that generates it.

Notice, however, that Y is scale dependent; if Y was originally dollars then going from a helicopter drop of dollars to one of cents for a given income level would increase Y and μ 100 fold. For fixed n it would also decrease the variance of income shares 100-fold. Both the dollar and cent drops are fair processes and there is no way a priori to choose the level of bundles being distributed. If this were the end of the story our contribution would be solely taxonomical; we would have shown that Theil's first inequality indicator (T_1) may be interpreted as a test of fairness under our fair (helicopter drop) income generation process. However, we can in fact use T_1 as a test statistic by estimating μ under the null. Below we show how to implement the test using a null consistent estimator of μ.

1.2. *The 'Helicopter Money Stop'*

We now consider a second income process. Suppose that, within any year, each of our n individuals receives the same amount of income each 'hour' and will continue to receive this hourly amount subject to a fixed hazard of exiting the income receipt process for the rest of the year, the 'helicopter money stop'. Under a fair null these hazards—the probability of 'stopping'—are the same each hour for each individual regardless of how many hours they have survived the hazard. Suppose further that we allow both the time interval and hazard probability to tend to zero. The result of this fair process is that each individual's income is a draw from an exponential pdf with the same mean β. This result maps back into the exponential distribution's natural role in modelling inter-arrival times. The income process is memory less in that if we see an individual with income x and who has survived the exit hazard their expected income will be $x + \beta$.

Consider the *LR* test of a null that y_i has pdf $EXP(\beta)$ versus the alternative that y_i has pdf $EXP(\beta_i)$. Under the null we have

$$\ln L(y_1, ..., y; \beta_0) = l_1 = -n \ln \beta_0 - \sum_{i=1}^{n} \frac{y_i}{\widehat{\beta_0}}. \tag{3}$$

The mle of β_0 is just average income (per head) $\overline{y}$. Under the alternative, we have

$$\ln L(y_1, ..., y; \beta_1, ..., \beta_n) = l_n = -\sum_{i=1}^{n} \ln \beta_i - \sum_{i=1}^{n} \frac{y_i}{\beta_i}.$$

The mle of β_i is just y_i so that l_n simplifies to

$$l_n = -\sum_{i=1}^{n} \ln y_i - n.$$

Straightforwardly then the LR test of $\beta_1 = \beta_2 ... = \beta_n$ is

$$LR_2 = 2(l_n - l_1) = 2n(\ln \overline{y} - \frac{\sum_{i=1}^{n} \ln y_i}{n}) = 2nT_2,$$

where T_2 is Theil's 'second' inequality index and, unlike the *LR* test in subsection 1.1, is scale free. Of course T_2 is the MLD, which is fast becoming the workhorse of applied inequality measurement.

As was the case with T_1, we cannot appeal to standard likelihood theory to ascertain the limiting distribution of *LR* in the helicopter stop process because the model under the alternative is saturated. However, implementing this and the helicopter drop test are quite easy, as we now show.

1.3. *Implementing the Tests*

The helicopter stop test statistic has a distribution that depends only on n so that for any given n approximate p-values may be computed via numerical simulation. For the multinomial we may estimate μ using the half mean (see below), and treat this estimate as the true null value of μ. Conditional on this estimate the null distribution now depends only on n and again its percentiles may be computed via numerical simulation.

Alternatively, the statistics may be linearly transformed to offer a large n normal approximation. In particular for LR_2 we show in the annex that

$$LR_2^* = \sqrt{n}\{T_2 - c\} = \sqrt{n}(T_2 - c) \to N(0, d),$$

where $c = 0.57785$ and $d = 0.64973$[9] and where we look to reject in the right hand tail only (we wish to reject for large values of T_2 not small).

The multinomial is scale dependent. To implement the test based on T_1 we will need to estimate μ. We do so under the null and in a way that fully exploits the null property of (asymptotically) normally distributed income shares. We specialise the fairness process to one where the number of income units per head μ is fixed but Y and n are large.[10] We use the fact that income shares for large n are approximately normal and estimate μ using the half mean. Explicitly in the annex we show that if $\widehat{\mu}$ is estimated as

$$\widehat{\mu} = 2(\pi \widetilde{s}^2)^{-1},$$

where $\widetilde{s}$ is the sample mean of $|ns_i - 1|$ then

$$LR_1^* = \sqrt{n}\{T_1 - b(\widehat{\mu})\} \to_D N(0, a(\mu)),$$

$$\text{where} \qquad b(\mu) = \sum_{j=1}^{\infty} \frac{(2j)!!}{\mu^j (2j)!},$$

where !! denotes double factorial, $' \to_D'$ means 'tends in distribution to' as $Y, n \to \infty$ with $\frac{Y}{n} = \mu$ and $a(\mu)$ is a variance term that depends only on μ. Given an estimate of $\widehat{\mu}$ we may consistently

[9] The constant c is just $c = \sum_{n=2}^{\infty} (-1)^n !n$ where $!n$ denotes subfactorial.

[10] We would also expect for any reasonable process that showers income units randomly on the population that μ be large and certainly exceed unity, because this quantity is just the average number of income packets per person.

estimate $a(\mu)$ numerically as follows. Draw k (large) *iid* samples of $z_i \sim N(0, \frac{1}{\widehat{\mu}})$ and use these as synthetic income shares to calculate k values of $\sqrt{n}\{T_1 - b(\widehat{\mu})\}$. Then use the sample variance of these k values as a consistent estimate of $a(\mu)$.

2. Testing for Inequality of Opportunity in Lotteries

In Section 1 we showed that both of Theil's inequality indices could be interpreted as test statistics of a null of 'fairness', meaning by this 'sameness' of lotteries faced by individuals within a class of income-generating processes. However, as noted in the Introduction, lotteries being different across individuals would not necessarily signify unfairness in the 'equality of opportunity' perspective of Roemer (1998) and the post-Roemer literature (see, for example, the review by Roemer and Trannoy, 2015). The question, rather, would be whether this variation was legitimate or illegitimate from a normative point of view, which in turn would be related to whether there was variation across exogenously given characteristics of individuals such as race. In this section we extend the analysis of Section 1 to test for sameness of lotteries across groups in society. We also relate our test statistics to the metrics of inequality of opportunity developed by Roemer and applied extensively in the post-Roemer literature.[11]

To assess the extent of illegitimate inequality Roemer presents a ratio of T_2 indices; the numerator is a measure of the extent of illegitimate inequality and the denominator of total inequality. This ratio has become the workhorse of a burgeoning empirical literature on the measurement of inequality of opportunity. We analyse this ratio now and show that it maps back into test statistics for fairness. But we also argue that this ratio is not a natural test of the null of a fair process; in fact, the spirit of our approach leads us to consider only the variation arising from the numerator of this ratio.

In what follows we assume that the population divides into two subgroups. We focus on two groups purely for notational simplicity and clarity: extension to k subgroups is trivial and briefly discussed below. We start with the helicopter stop process null and its corresponding test statistic T_2.

Suppose that there are $n_1(n_2)$ individuals drawn from two mutually exclusive and exhaustive subgroups of a population. We will assume that the sample is ordered so that group one observations appear first. It is well known that Theil's second inequality index T_2—the MLD—may be decomposed into a within group index plus a between group index as follows

$$\begin{aligned} T_2 &= \{p_1(\log \overline{y}_1 - \overline{\log y_1}) + p_2(\log \overline{y}_2 - \overline{\log y_2})\} \\ &\quad + \{\log \overline{y} - (p_1 \log \overline{y}_1 + p_2 \log \overline{y}_2)\} \\ &= \{T_2^L\} + \{T_2^I\}, \end{aligned}$$

where $\overline{y}_1 = \frac{\sum_{i=1}^{n_1} y_i}{n_1}$, $\overline{y}_2 = \frac{\sum_{i=n_1+1}^{n} y_i}{n_2}$ and $p_i = \frac{n_i}{n}$, $i = 1, 2$.

The within group component T_2^L is the weighted sum of the T_2 indices computed from each subgroup separately. In Roemer's formulation it represents the 'legitimate' variation in the Theil index—the income variation that is a result of personal effort rather than inherited circumstance. The between component T_2^I is the Theil index computed assuming that all individuals within a subgroup have income equal to the group's average In Roemer's formulation T_2^I measures

[11] See the survey by Roemer and Trannoy (2015).

the ‘illegitimate’ component of the Theil index—the income variation that is a result purely of inherited circumstances.

Using the results in the previous sections we can readily see that T_2^I is merely $2n$ times the LR test of the helicopter stop rule null (equal mean stopping times for all individuals regardless of subgroup) against the alternative that the two subgroups have different mean stopping times, i.e.,

$$2nT_2^I = l_2 - l_1,$$

$$\text{where } \; l_2 = \max_{\beta_1\beta_2}\{-n_1 \ln \beta_1 - n_2 \ln \beta_2 - \sum_{i=1}^{n_1} \frac{y_i}{\beta_1} - \sum_{i=1}^{n_2} \frac{y_i}{\beta_2}\}$$

and where l_1 is as defined in (3). Once more, we have a different and concrete interpretation of a Theil index as a test statistic of a null of a fair income process against an alternative that circumstances outside the individual’s control affect this process. The statistic is a χ_1^2 under the null, so instead of presenting the numerical value of the index one may present its corresponding p-value. When comparing illegitimate Theil indices from two populations we could then compare p-values rather than Theil index levels.

However, as we have noted, Roemer uses a ratio of illegitimate to total variation as a measure of inequality of opportunity. This share—R_2, say—can be written as

$$R_2 = \frac{T_2^I}{T_2} = \frac{\ln \overline{y} - p_1 \ln \overline{y}_1 - p_2 \ln \overline{y}_2}{\ln \overline{y} - \overline{\ln y}},$$

where T_2^I is the illegitimate between group Theil index. Given the decomposition above, this has a clear-cut interpretation: it is the proportion of income variation (as measured by Theil) attributable to illegitimate inequality. Driving our approach to its logical conclusion, we could also interpret R_2 as a ratio of LR tests with the same null of fairness as per our helicopter stop process. Explicitly, a simple application of the analysis in Section 1 shows that

$$R_2 = \frac{2nT_2^I}{2nT_2} = \frac{LR_2^I}{LR_2},$$

where LR_2 tests the fair helicopter stop rule null against an alternative that individuals have different means and LR_2^I is the LR test of the same null against the alternative that each subgroup has its own distinct mean (each individual draws from her particular subgroup’s exponential pdf).

As we have already indicated, using the variation in numerator and denominator for a test runs against the spirit of our approach. Even if we could deduce the null distribution of R_2, the test would have power properties that would make rejections and failures-to-reject hard to interpret. In particular, the power of the test against illegitimate inequality is moderated by the extent of legitimate variation. Instead, we suggest modifying the ratio to use only the variation in the numerator of R_2.[12] Explicitly, note that nR_2 can be written as

$$nR_2 = \frac{2nT_2^I}{2T_2} = \frac{n(\ln \overline{y} - p_1 \ln \overline{y}_1 - p_2 \ln \overline{y}_2)}{\ln \overline{y} - \overline{\ln y}} = \frac{\chi_1^2}{c} + o(1),$$

where $c(= 0.57785)$ is the null limit of T_2 as defined above.

[12] A referee has also pointed out that the Roemer ratio would fall if a change in the way effort was rewarded led to an increase in within-group inequality, even if between-group inequality remains the same. This is, of course, an issue for the normative framing of the Roemer measure and its application in the literature.

This statistic is a standard LR test. It follows a χ_1^2 distribution. It is this statistic—extended to the case of k groups—that we compute in our empirical work in Section 4.

We now turn to consider the helicopter drop process (where Theil's first inequality index T_1 is the relevant data quantity). Following previous logic, we could write T_1 as (proportional to) the sum of two LR tests as follows

$$2n\mu T_1 = 2\{l_2 - l_1\} + 2\{l_n - l_2\} = \{2n\mu T_1^I\} + \{2n\mu(T_1 - T_1^I)\}, \tag{4}$$
$$\text{where } T_1^I = \sum S_i \ln(\rho_i . S_i),$$

where S_i is now group i's share of the total income Y and where $\rho_i = n/n_i$ The first statistic in (4)—$2n\mu T_1^I$—tests a null of equal means against an alternative each group has its own mean. If we treat all of the people in each group as a single person then we see it is a measure similar to T_1 the difference being the replacing of n in T_1 with ρ_i. The second statistic in (4) tests the null that each subgroup has its own mean versus an alternative that each individual has her own separate mean. If we accept that some proportion of income inequality is due to 'legitimate' reasons then our approach suggests we should focus on the first component only. Analogous to nR_2 above, we propose to compute the ratio nR_1 where

$$nR_1 = \frac{2n\mu T_1^I}{2\mu T_1} = \frac{nT_1^I}{T_1} = \frac{\chi_1^2}{const} + o(1),$$
$$\text{where } const = 2\mu . plim T_1 = 2\mu b(\mu),$$

where $b(\mu)$ is as defined above. Unlike before, we do not need to estimate μ as long as it is larger than unity. The parameter μ is the average number of income 'packets' that an individual receives. In our sample below, average income is around \$15,000, so a value of unity for μ would imply that the helicopter is dropping packets of this size. While technically this is fair, by our reasoning such large packets would certainly guarantee—ex ante—a very high variation in incomes after the helicopter drop has occurred. A sensible prior view therefore would be that μ is larger than one. Finally on this point, in the empirics below our estimate of μ (under the null) is around 2.5. We therefore use the fact that $\mu b(\mu) \approx \frac{1}{2}$ to compute (an approximation to) nR_1 as

$$nR_1^* \approx \frac{nT_1^I}{T_1}$$

and compare it to critical values of the χ_1^2 distribution. Finally, note that because $T_1 - T_1^I > 0$ R_1 is always between zero and one the decomposition of T_1 into T_1^I and a remainder term allows us to interpret R_1 as the proportion of T_1 that results from illegitimate income variation.

Extending the above to k mutually exclusive and exhaustive groups (for example, white and male, white and female, non-white and male, non-white and female) is trivial. Under the alternative the log likelihood is 'additively separable'; the k parameters appear separately in k separate additive terms. The formulae presented above for the helicopter stop rule remain intact and unchanged. For the helicopter drop (4) we must replace the number 2 with k in the formula. Of course, in this case the resulting tests are chi-squred statistics with $k - 1$ rather than 1 degrees of freedom.

We close by noting that the statistics above are analogous to the nR^2 F-test of significance of the regression of income on group dummies. Under the null of no illegitimate variation this regression has no explanatory power while under the unfair alternative it does.

3. A Bayesian Perspective

Until now, we have motivated use of Theil's indices using a purely classical view that we wish to make a 'decision' as to whether our income process is fair. An alternative motivation for their use comes through their relationships with Bayes' factor, as we now show.

If we treat the 'fair' and 'unfair' processes (null and alternatives above) as models and if we assign equal prior probability to their respective 'truth' we may interpret the LR in terms of the ratio of posterior probabilities of fair to unfair models. In particular Kass and Raftery (1995) show that for large n the log of the Bayes factor (log(BF)) is approximately

$$\log(BF) \approx LR - k.\log(n)$$

where LR is the difference in the log likelihoods between unfair and fair models and k is the number of extra parameters moving from the fair to unfair models.

There are several things about the use of log(BF) that make it attractive in our context, particularly when we wish to compute Theil indices from different datasets with possibly different sample sizes (n). First, it is not a decision-making tool. Instead—as we have noted above—it gives us the posterior probability that the process is unfair divided by the posterior probability it is fair.[13] It does not therefore require that one or other of our 'models' be true—it just offers a measure of the models' relative concordance with the data. By contrast the p-values of the classical tests (above) are valid as probability statements only when the null is absolutely true. Additionally, even if the estimated parameters of the processes were close to one another in the two models in a quantitative sense, the p-value will still tend to zero with n. Hence the p-value may mask this quantitative closeness. An additional problem is that comparisons of results are difficult across different datasets if they have different numbers of datapoints. We may, of course, solve this problem by simply bootstrapping the data and computing estimated p-values over a range of n.[14] However this procedure is computationally inconvenient. A further—admittedly minor—issue associated with large n is that p-values may be too small to compute accurately.

Set against these arguments in favour of the BF is the fact that, to compute it for our helicopter drop process, we need an estimate of the scaling parameter μ, and this is only identified under the null. While this (nuisance) parameter takes the same value in both models, the estimate that we obtain is only consistent under the null. By contrast, nR_1^* may be computed without an estimate of μ, as can nR_2 (which has no scaling parameter). Therefore we only compute the BF's for the analysis of illegitimate income variation.

4. Empirical Illustration

To provide an illustrative application of the above tests, we drew a sample of incomes from the 1979 National Longitudinal Survey of Youth (NLSY) for the year 1995. The 1979 NLSY is a nationally representative sample of people who were between the ages of 14 and 22 in 1979, so, by the year 1995, nearly all members of the sample will be young adults who have completed their education. As always in a single cross-sectional snapshot, we are using current income as a proxy for permanent income, which is obviously flawed. In particular, many people in our

[13] Making the posterior probability of the alternative the numerator and that of the null the denominator is a convention that we adopt to match the spirit of the LR test.

[14] A bootstrap based simply on random resampling of the data is valid under our null hypothesis of fairness, because, under the null, incomes are indeed independent of one another.

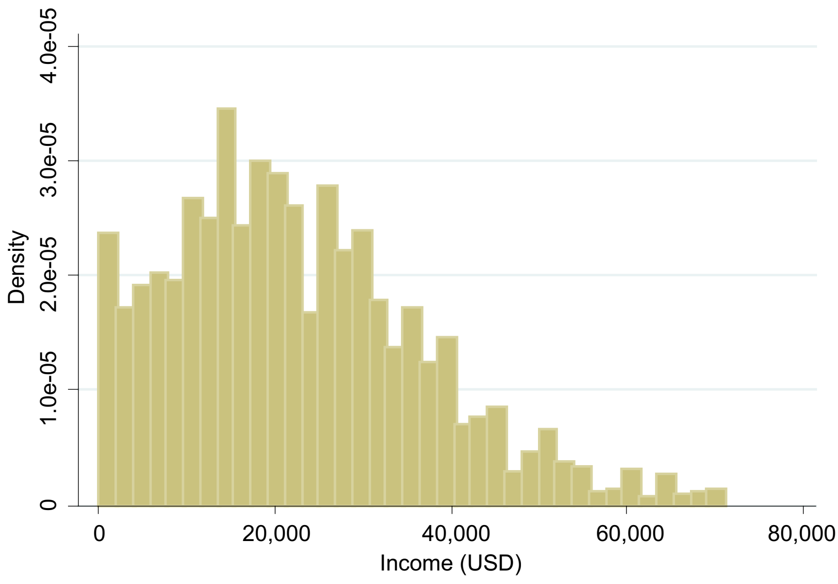

Figure 1. *Raw Income Distribution.*

sample recorded zero income for the year. Either they are not in the formal workforce (e.g., married workers in the home) or they are unemployed. Given that our exercise is only illustrative, we do not attempt to adjust income of individuals to take account of the disconnect between permanent and current income, nor do we try to estimate household income. We simply drop from the sample those people who recorded zero income. After dropping data we were left with 5,736 individuals.

We compute the four test statistics LR_1^*, LR_2^*, nR_1^* and nR_2. To compute the last two of these—based on the illegitimate variation in incomes—we control for race (white/non-white), gender, region of upbringing (south/non-south), whether the individual has a work-hindering health condition and whether the child was raised by biological parents. This gives ($k =$) 32 mutually exclusive and exhaustive groups of people. Figure 1 shows the distribution of raw incomes (total variation), while Figure 2 shows the distribution when each individual is assumed to earn their group mean income (illegitimate variation). The graphs are radically different. The empirical pdf of raw incomes in Figure 1 shows some resemblance to an exponential pattern and also resembles a left truncated normal pdf. Figure 2, on the other hand, shows a distribution of illegitimate income that resembles neither exponential nor normal.

Table 1 presents the four test values and their p-values. We see that LR_1^* (helicopter drop process) and LR_2^* (helicopter stop process) are negative, implying that the likelihood ratios (which are always strictly positive) are less than their respective expected values under the null. Our rejection region is always in the right tail—only higher, and not lower, than expected Theils can reject.

The 'illegitimacy' statistics for the two processes nR_1^* and nR_2 both resoundingly reject their respective nulls with near infinitesimal p-values. As we have already noted, small p-values are the consequence of large samples; with over 5,000 observations, even local departures from the null will result in very large test statistics.

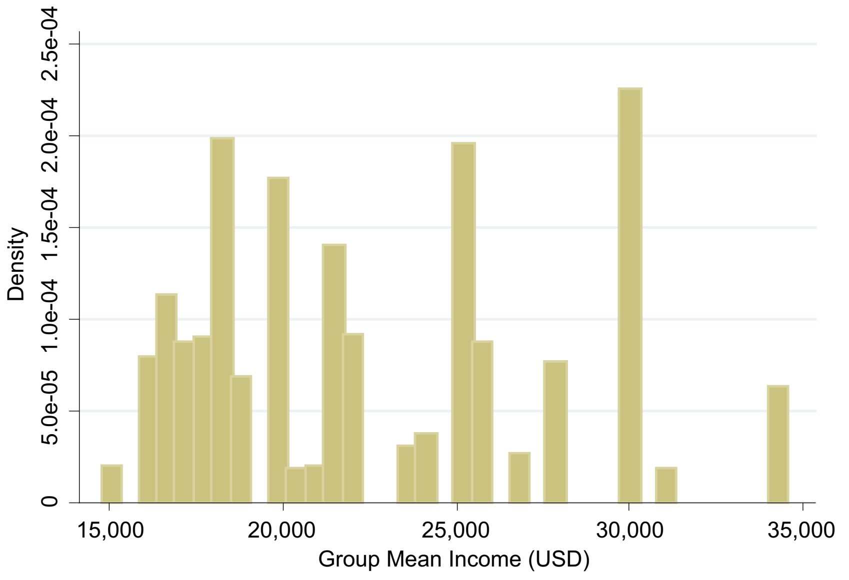

Figure 2. *Distribution of Income at Group Means.*

Table 1. *Tests and p-values.*[15]

Statistic	LR_1^*	LR_2^*	nR_1^*	nR_2	BF_1	BF_2
	−3.55	−22.76	660.44	334.75	383.49	55.56
	−0.0005	-10^{-115}	−3,047	−727		
$\log(p - value)$						

Notes: LR_i^* are standard normal and nR_1^* and nR_2 are χ^2_{32} variates.

Finally Table 1 also shows the two Bayes factors for fairness versus illegitimate variation. Here we see a reinforcement of the corresponding test results for nR_1^* and nR_2: the unfair alternative of illegitimate variation is very much more likely in posterior terms than the null of legitimate variation.[16]

5. Adjusting the Tests for Heteregeneous Effort

Returning to the framework of Section 1, our test for fairness is that individuals face the same income-generation process, within the overall class specified (helicopter drop or helicopter stop). But non-identical income prospects do not necessarily imply unfairness. If, through their effort, individuals can transform the basic lottery of 'drop or stop' into even greater income-earning

[15] We approximate the p-values using $p = \frac{f(x)^2}{2f'(x)}$. This is the area of the triangle whose height is the height of the pdf at the statistic's value x, and whose hypoteneuse is the tangent to the pdf at x. It is an underestimate.

[16] Classical tests and Bayes factors are conceptually and statistically completely different data measures, and as a result do not always point in the same direction vis à vis their relative support for the null and alternative. For example, in their seminal paper on Bayes factors Kass and Raftery (1995) note 'a dramatic example with $n = 113{,}566$ ... A substantively meaningful model that explained 99.7% of the deviance was rejected with a standard chi-squared test with a p-value of about 10^{-120} but was nevertheless favored by the Bayes factor' (op cit. p. 789). In other empirical work, not reported here, we have indeed found a minuscule p-value for the classical test 'overturned' by a negative BF.

prospects, surely they deserve these prospects, and variation in effort must be taken into account in testing for fairness. The question we now ask is: how, if at all, may we adjust our test statistics to allow for the existence of heterogeneous effort in the population? There is no general way of doing this, but some progress is possible if we make some assumptions about how effort impacts on income.

Suppose that we start with an economy in which effort is homogenous and individuals draw income from some pdf (which may or may not be the same, depending on whether we are operating under the null or the alternative). We then move to an economy in which total income (and hence mean income) is the same as before[17] but where effort varies across individuals; in effect, we are then considering the impact of heterogeneous *relative* effort rather than *absolute* effort. If we assume that relative effort is (a) independent of the individual's income draw they face (b) multiplicatively scales up/down base income, (c) has a mean of unity and (d) is uniformly distributed for all individuals, then we may make simple numerical adjustments to our test critical values and proceed exactly as before.

Under these assumptions, we may write observed income for individual i y_i as (unobservable) base income y_i^* times (unobservable) effort e_i. The log likelihood of observed income for individual is then

$$\begin{aligned}\ln L(y_i, e_i; \theta_i) &= \ln L(y_i|e_i; \theta_i) + \ln L(e_i)\\ &= \ln(\frac{1}{e_i}) + \ln L(y_i^* e_i|e_i; \theta_i) + \ln L(e_i).\end{aligned}$$

For any base income pdf the first and last terms will drop out of the LR test of fairness ($\theta_i = \theta\ \forall i.$). If we now repeat the LR tests for the pdfs in Section 2, the functional forms remain as in Table 1 but the null pdf of the y_i is now the product of a uniform effort variate and the base income variate. To illustrate, consider the LR test for the exponential case. This has the same formula as before, except that now we use $y_i^* e_i$ in place of y_i—that is:[18]

$$\frac{LR_2}{2n} = \ln \overline{y^* e} - \overline{\ln y^* e}.$$

Now, instead of the index being made of exponentially distributed variates, it contains the product of exponential and uniform variates. The uniform has known support—namely [0, 2]—and the RHS is invariant to scale so may be treated as an exponential with scaling parameter one. Consequently, once we have computed LR_2 from the data we may generate its corresponding p-value straightforwardly via simulation. To illustrate how effort changes the null pdf of the test statistic, we simulated the distribution of T_2 for the regular (no effort) case for $n = 5{,}000$ (see Figure 3) and compared it with the current case with heteregeneous effort (Figure 4). As expected, the effect of heterogeneous effort on the fair distribution is to increase its spread and shift it by about 0.4 to the right. In the context of our fairness test, allowing for heterogeneous effort will be 'more forgiving', in the sense that, for any given MLD index value, the p-value will be considerably larger than under assumed homogenous effort.

The assumption that all agents draw effort from the same pdf is reasonable under the null, but rather more controversial under the alternative. Under the null, ex ante every agent faces the

[17] The assumption of same total (and mean) income is a normalisation. Inequality indices measure relative incomes and are invariant to scaling up or down of total income. In our set up relative incomes will only change when when relative effort changes.

[18] We may also use the independence of effort and the realisation of income to approximate $LR_2/2n$ as $\ln \overline{y^*} - \overline{\ln y^*} + \ln \overline{e} - \overline{\ln e}$.

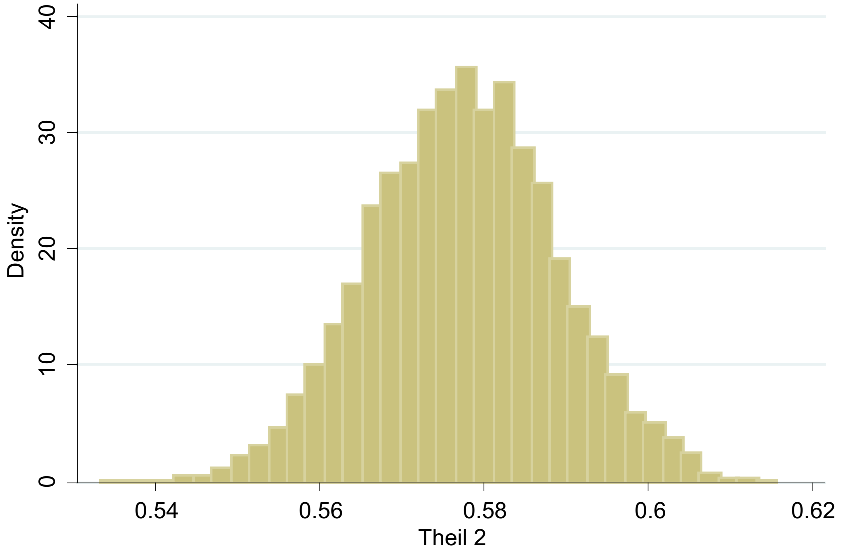

Figure 3. *Theil 2 (MLD) Under Homogenous Effort.*

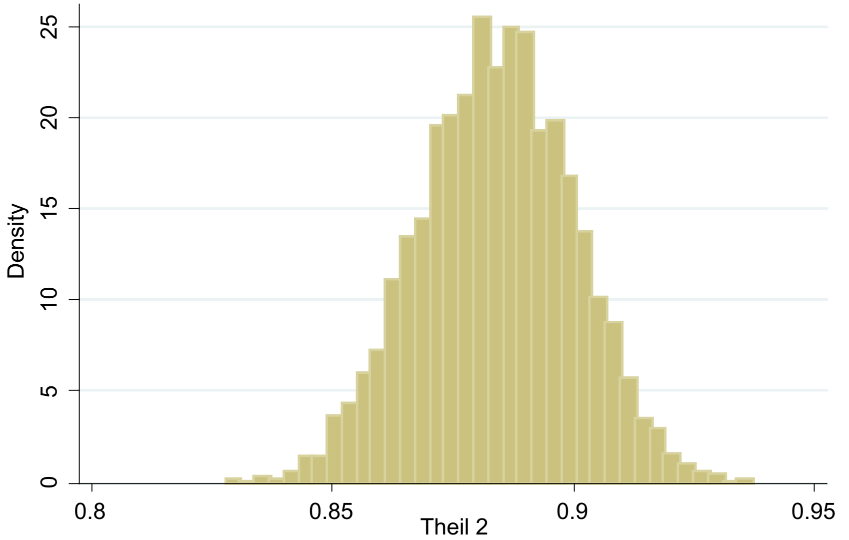

Figure 4. *Theil 2 (MLD) Under Heteregenous Effort.*

same pdf for pre-effort or 'base' income, so there are no differential incentives to exert more or less relative effort during the income-generating process. Under the alternative, however, those agents facing pdfs for base income with higher means may make more effort, because the rewards of doing so are greater than the average. In our defence we could argue—as is often done in macroeconomics—that the income effect of facing a higher mean base income cancels out its substitution effect.[19]

In the context of the education accrual model discussed in the next section, extra relative effort will decrease the probability of 'stopping' in any time period (failing the end period exams that allow progression to the next period) and hence increase the overall mean time spent in education. In the context of multinomially distributed educational units, extra relative effort will increase the chance that the individual receives each particular unit.

Assumption (c) reflects the fact that we wish to measure relative effort and is merely a normalisation. The assumption that effort is uniformly distributed across the population seems quite strong. However, if we are completely ignorant about the process generating effort then, from a Bayesian perspective, the uniform distribution (a 'diffuse prior' in Bayesian terminology) is a natural choice. The strongest assumption of all is (b). The main justification for this assumption is that it allows us to make definitive adjustments to our tests/indices, whereas without it we could say nothing about the impact of effort on our tests. We should note that the assumptions (c) and (d) imply that, in a fair economy, the maximum impact that relative effort may have on an individual's income is to double it. In fact this is true under our alternative also, but here the empirical content of the statement is empty because we do not know (and cannot estimate) what the individual's income pdf is.

6. Interpreting the Lotteries[20]

Our approach in this article has been reduced form in nature. We have presented two canonical families of lotteries and offered a specification of fairness within each class. In the helicopter drop process it is equality of the probability that a dollar from the helicopter accrues to any individual. In the helicopter stop process it is equality of the probability that an individual is forced to exit the income receiving flow. We believe that as reduced form representations these characterisations have considerable intuitive appeal. In essence, our two processes highlight mechanisms whereby individuals can have different probabilities of earning income, or different probabilities of stopping earning income. We now discuss a number of structural economic processes that might lead to outcomes approximated by these reduced forms.

One route to providing an economic interpretation of such mechanisms is through processes of schooling and human capital formation. A feature of education is that the taking of an incremental year of it usually requires success during the previous year; in the UK, students typically need General Certificate of Secondary Education (GCSE) passes (high school exams for 16 year olds) in order to take A levels (high school exams for 18 year olds), while people will typically need to have passed A levels in order to progress to a university degree, and so on. In essence, then, we might think of the total amount of education attained as a stopping process; during each unit of time in education there is a chance that the individual may go no further and that his or her education will stop. If all individuals have the same stopping hazard then the process is fair.

This story has extra resonance when we allow for heterogeneous effort, as per the extension in the previous section; it is natural to think of those who study harder as having higher chances of

[19] An alternative justification would be if effort we made 'upfront', i.e., before the lottery became known to the individual. In the context of effort during the acquisition of education (one of the economic mechanisms that we advance to support the helicopter stop process below) assuming that effort is invested before the lottery occurs is quite natural.

[20] We are grateful to a referee for detailed comments and questions on this topic—we have drawn heavily on their report to formulate this section.

continuation up the education/human capital ladder. If units (years) of education have a constant market price then the distribution of education, and thus the distribution of incomes, will reflect the differential probabilities of exiting the education process. Of course, this economic process implies that human capital earns rents that cannot be competed away. However, the existence of rents is only unfair if, ex ante, individuals have different chances of earning them.

As an illustration of how this stopping process for human capital might work in a general equilibrium model that has physical capital as well, in the Annex we embed the process into a standard Diamond overlapping generations (OLG) growth model. There we allow the level of accrued human capital to factor up the number of efficiency units of the individual. Under the null of equal stopping probabilities, cross-sections of incomes of a particular age group are approximately exponentially distributed.[21]

Staying with the stop process, another interpretation is related to unemployment shocks. There is evidence that spells of unemployment increase the probability of further unemployment.[22] There is hysteresis in the duration of unemployment spells. Our stop process, while not exactly the same, captures the idea of unemployment as being the significant slowing down of income flow. If the probabilities of unemployment were identical across individuals, this income-generation process might be considered to be fair. The actual income distribution will reflect the differential probabilities of unemployment and reduced income flow.

Coming now to the drop process, one simple interpretation is in terms of monetary expansion (recall that the image often used for this is ‘helicopter money’). Although this applies to increments of income rather than total income, the basic idea would be that new injections should be distributed fairly across the population rather than helping one part of the economy—for example, those owning stocks whose values become inflated. A similar argument can be made for increases in public expenditure: fairness might require that they should not benefit (in probability) one section of the population more than another. A specific case in point would be government expenditure on technological research and development. If this ends up benefiting certain groups more than others then the observed distribution of income will reflect this unfairness, which could, in principle, be tested for.

The large literature on income/earning dynamics also speaks to providing economic rationales for the ‘stop’ and ‘drop’ processes. The ‘drop’ process leads to multinomial outcomes, which are approximately normally distributed. If incomes are the subject of cumulated idiosyncratic iid shocks at each point in time[23]—shocks from any pdf—then incomes will be approximately normally distributed, consistent with the multinomial process for a large economy. Blundell (2014) outlines a model of income inequality over the life cycle, in which individual incomes are subject to idiosyncratic stochastic trends. Key to this model is the presence of constraints on household borrowing and income insurance. A special case of that model— where the innovations at time t for cohort i are drawn from the same pdf across individuals—could be considered to be fair. In this null, and where the idiosyncratic trend had been running for some time (for older workers, say), incomes would be approximately normally distributed in the cross-section.

Thus the ‘drop’ process would determine how earnings develop as the effect of idiosyncratic shocks. Under the null, these shocks hit every worker with equal ex-ante intensity. As discussed

[21] The stopping rule yields human capital that has infinite support, whereas if we wish to map this process into real time the amount of human capital will be truncated in its right tail. Consequently, the distribution of incomes are only approximately exponential.

[22] See, for example, the well-cited papers of Huff Stevens (1997), Lancaster (1979) and Nickell (1979).

[23] The pdf of shocks may be different each time period, but under a fair null all individuals would draw shocks from the same pdf at time t.

above, the 'stop' process might now refer to the way in which unemployment probabilities (i.e., the probability of exiting from the earning process within the relevant unit of measured time) are distributed across the population. Under the null, one expects everybody to have ex ante equal chances of being unemployed, hence the same expected earnings. It should be noted, however, that the processes only lead to their respective distributed outcomes under the null of fairness. The distribution of outcomes under the alternative is much more complicated to analyse. In particular, it would be analytically intractable in our framework to compound the 'drop' and the 'stop' processes to derive a single test statistic for joint fairness in both.

7. Discussion and Conclusion: Fairness and Testing

What is 'fair'? A coin is said to be fair if it has equal probability of coming up heads or tails. A fair toss is the toss of a fair coin. If all outcomes were decided by the toss of a fair coin, the outcomes would be fair in this sense. More generally, the key is sameness of prospects across all individuals. This brings us to 'equality of opportunity', which might be captured by saying that prospects should not depend on factors outside individual control. Thus, simply measuring inequality of outcomes is not a way of gauging inequality of opportunity or fairness. And inequality of outcomes is particularly problematic when luck is involved, because outcomes can be unequal even if life chances were identical before the luck of the draw. However, the question is whether observed inequality can nevertheless provide information on whether the chances were or were not the same across individuals. We argue in this article that it can.

Thus our article contributes to the literature on inequality, fairness and equality of opportunity by bringing into sharp relief and relating two key features. The first is the idea that observed outcomes are dependent on luck. The second is that if we put sufficient structure on the nature of this luck, we can continue to use conventional inequality measures, but now as test statistics for fairness. We have shown how, for two canonical processes, Theil's two measures can be used as test statistics for fairness in the sense of all individuals facing the same lottery of life. We have further extend this to testing for whether lotteries differ across groups with exogenously given characteristics, and linked this to the Roemer measure of inequality of opportunity. As an extension, we have provided a Bayesian perspective on these tests. We have also shown how, under certain assumptions, the form of the test can be preserved when effort modifies the basic lottery in the income-generation process. While we believe that our reduced form approach has considerable appeal, we have also discussed conditions under which economic processes could approximate to these reduced forms.

We have, of course, only scratched the surface of the vast literature on fairness and equality of opportunity. It is true that fairness in the process is important, and that equality of opportunity is a major concern for social justice. However, many people also care about ex post equity—that is, equality in outcomes. There is a whole literature on inequality and fairness in risky situations (or for risky income-generating processes such as those described in the article) stemming from Diamond (1967), who promotes fairness in the process, and Broome (1984) who promotes fairness in outcomes.[24] Thus, this literature claims that both ex ante and ex post inequality matter and should be accounted for.[25] Fleurbaey (2018) provides an up-to-date review of the welfare economics of risk and uncertainty and shows the remarkable extent to which the literature is

[24] We are grateful to a referee for raising this point. We use the referee'a report to formulate our discussion.

[25] See, for instance, Ben Porath *et al.* (1997) and Saito (2013)

still grappling with these issues. Furthermore the notion of 'luck' itself is much contested, as discussed and developed in Lefranc and Trannoy (2017) Also, in the inequality of opportunity literature the principle of compensation (that different groups or individuals should face the same prospects or opportunities) is complemented by a principle of reward (how the outcome should be distributed between people who faced the same opportunities). The approach in this article is consistent with the 'liberal' reward, that any inequality in income arising from the income-generating process is unproblematic (provided that everyone faced the same income-generating process). But it is clearly problematic and has been questioned in the literature. Kanbur (1987) argues that, with extreme outcomes, inequality of the ex post distribution may dominate ex ante equality of a fair coin toss, and Kanbur and Wagstaff (2015, 2016) offer an extended critique of the Roemer and post-Roemer measures of inequality of opportunity, arguing for greater weight to be given to the actual outcomes.

In the axiomatic literature,[26] our approach perhaps comes closest to that of Mariotti and Veneziani (2017), who propose 'formulating opportunities as chances of success': 'To model opportunities, we assume that each individual is regarded as a binary experiment with either "success" or "failure" as possible outcomes. Then, opportunities in society are expressed by the role of chances of success across individuals.' (Mariotti and Veneziani, 2018). The focus of their paper is somewhat different to ours, of course. They set out axioms for liberal justice and rationality, and conclude that there are 'strong limits to inequality in the profile of opportunities that are implied by a liberal principle of justice and of social rationality'. However, in our context, think of each individual facing a lottery of life with a probability of success or failure. Identical lotteries would characterise fairness in our setting. The question for us is: how would we know that the lotteries were, in fact, identical, when all we can observe are the outcomes? This leads us to testing for fairness.

Testing for fairness in the presence of luck presents a set of statistical issues that go beyond those discussed in the literature to date on equality of opportunity. A starting point is to ask whether lotteries are the same or different across individuals. A rejection of the hypothesis that lotteries are the same across individuals is a necessary (but not, as discussed at length in the article, a sufficient) condition for unfairness. The most general specification for fairness is to envisage each individual's income as an independent draw from the same single but unspecified/unknown pdf. It is easy to see that testing this generic fair null is impossible using a single cross-section of incomes; for any set of income realisations we can always find a pdf for which the null is not rejected and a pdf for which it is. For example—and trivially—we could never reject the null that incomes were independent draws from the observed empirical pdf itself! To gain traction from a single cross-section there must be a specific class of income-generating processes that is the frame for testing the null that the process is the same for each individual.

Thus the test for fairness in a stochastic setting must be conditional on the class of income-generating processes that frame that test. The two processes that we discuss in this article have, we think, considerable intuitive appeal. We have also discussed the extent to which they could represent the reduced form of economic processes, such as dropouts from education or idiosyncratic shocks over the life cycle. But we acknowledge that economic processes could lead to other reduced forms for lotteries facing individuals. For each family of lotteries and each specification of fairness within that family, there will be a corresponding LR test statistic. Thus, for example, it is easy to show that the LR test statistic for fairness for draws from a family of

[26] The literature is, of course, huge. The book by Fleurbaey and Maniquet (2012) provides an overview up to that date; Roemer and Trannoy (2015) is a more recent review.

single parameter uniform pdf's $U(0, \theta)$ is given by:

$$LR_3 = \overline{\ln(y/y_{max})},$$

where y_{max} is the largest income in the sample and denotes sample mean. Similarly, the LR statistic for a test of the null that the draws were from a Pareto density with the same Pareto exponent is[27]

$$LR_4 = \ln \overline{y^*} - \overline{\ln y^*},$$

where $y^* = \ln y$.

LR_3 and LR_4 have something of a resemblance to LR_1 and LR_2, which are simple transforms of classical Theil inequality indices T_1 and T_2. In particular Theil's second inequality index T_2 is nothing other than the MLD, which is increasingly widely used in the applied inequality measurement literature. MLD is simply the log of mean income minus the mean of log income. But LR_4 is simply $2n$ times a log version of MLD—log of the mean of log income minus the mean of the log of log income.

We thus have a framework for generating test statistics for fairness upon specification of the family of processes and delineation of the null of fairness within that family. Of course, each test is only valid under the specific process that motivates its use. We believe that this is analogous to the machinery in inequality measurement for generating inequality measures based on specific social welfare functions, as proposed in Atkinson's (1970) classic contribution. Atkinson based his family of inequality measures on the concept of equally distributed income Y_{ede}—that level of mean income which, if distributed equally, would give the same level of social welfare as the current unequal distribution. The Atkinson measure of inequality is then given as:

$$A = 1 - (Y_{ede}/\mu),$$

where μ is the current mean income. This family of inequality measures depends on the specific social welfare function and the specific income distribution. For example, suppose that the social welfare function is the sum of log incomes:

$$W = \sum_{i=1}^{n} \ln y_i.$$

Then, it follows that:

$$Y_{ede}/\mu = e^{-T_2}.$$

Hence the Atkinson measure of inequality on the ex post distribution is a simple transform of the T_2 measure.

For any given pair of social welfare functions and any given distribution, the Atkinson machinery generates an inequality measure consistent with his perspective of social welfare and inequality. Similarly, for us, any given family of income-generating processes and a specification of fairness within that family generates a test of fairness. In some cases these test statistics are transforms of or bear a close resemblance to conventional inequality indices. However, we would suggest that the literature consider these statistics as possible inequality measures in their own right. Just as Theil (1967) appealed to entropy theory, and to Shannon (1948) and information

[27] We also assume that the unknown level of minimum income is the same under the null and the alternative.

theory, as a starting point to generate his inequality indices, testing for fairness could be starting point for generating new families of inequality indices.

Appendix: The Asymptotic Distributions of LR_1^* and LR_2^*

The multinomial LR test (T_1^*)*:-*

For the multinomial the phrase 'asymptotically' is taken to mean n, $Y \to \infty$ with μ fixed. Under the null and asymptotically, $ns_i \sim N(1, \frac{1}{\mu})$. Now consider the quantity $|ns_i - 1|$ which obviously has the half normal distribution. The mean of this half normal is equal to $\sqrt{\frac{2}{\pi}}$ times the standard deviation of ns_i which in turn is just $\frac{1}{\sqrt{\mu}}$. Hence a ($\sqrt{n}$) consistent estimate of μ is given by

$$\widehat{\mu} = 2(\pi\widetilde{s}^2)^{-1}$$

which is the formula given in the text. We use $\widehat{\mu}$ to estimate $b(\mu)$ in order to implement the test. Note that

$$\sqrt{n}\{T_1 - b(\widehat{\mu})\} = \sqrt{n}\{T_1 - b(\mu)\} + \sqrt{n}\{b(\mu) - b(\widehat{\mu})\} =$$
$$\sqrt{n}\{T_1 - b(\mu)\} + \sqrt{n}\{\frac{1}{\widehat{\mu}} - \frac{1}{\mu}).b^{'}(\widehat{\mu}) + o(1),$$

where $'$ denotes derivative with respect to $\frac{1}{\widehat{\mu}}$. The random variable $\frac{1}{\widehat{\mu}}$ is the square of the sum of *iid* variates and so $\frac{1}{\widehat{\mu}} - \frac{1}{\mu}$ is an asymptotically normal mean zero variate. However T_1 is also asymptotically normal as we now show.

Define $s_i^* = ns$ then we have

$$T_1 = \frac{\sum_{i=1}^n s_i^* \ln(s_i^*)}{n}.$$

This shows that T_1 is the sum of n *iid* variates each with finite variance and so is asymptotically normal. Now we find its large sample mean. Expand each $s_i^* \ln(s_i^*)$ element in a Taylor series around $s_i^* = 1$ (the null mean of s_i^*) to get

$$T_1 = 0 + \frac{\sum_{i=1}^n (s_i^* - 1)}{n} + \sum_{j=2}^{\infty}\sum_{i=1}^{n} \frac{(s_i^* - 1)^j}{j!n}$$
$$= \sum_{j=1}^{\infty} \frac{\sigma^{2j}}{(2j)!} + o(1) = b(\mu),$$
$$\text{where } \sigma^{2j} = \frac{(2(j-1))!!}{\mu^j}.$$

The last equality follows from the asymptotic normality of shares. In the last sum, '!!' stands for double factorial.

The exponential LR test (T_2^*)*:-*

First of all note that under the null the scale parameter of the exponential (the inverse of its mean) washes out of T_2. Hence, expanding T_2 in a Taylor series around $y_i = 1$ gives

$$T_2 = \sum_{j=2}^{\infty} (-1)^j \frac{\widehat{\sigma(j)}}{j!},$$

where $\widehat{\sigma(j)} = \frac{\sum_{i=1}^{n}(y_i - 1)^j}{n}$ is the jth sample central (around 1) moment of income. All sample moments are consistent for their theoretical counterparts and the weights are geometrically declining. So to get the *plim* of this expression we replace the sample moments with actual moments of the exponential's pdf with unit scale parameter. This gives an expression for the term c in the text as

$$c = \sum_{n=2}^{\infty} (-1)^n \, !n$$

$$\text{where } !n \; denotes \; n \; subfactorial.$$

Asymptotic normality of $\sqrt{n}(T_2 - c)$ is established by noting that each sample moment is the sum of n *iid* variates so each sample moment is asymptotically normal. We computed c—the variance of the statistic—by drawing k random samples of $y_1, y_2, ..., y_n$ from an exponential with unit scale parameter and used these y values as synthetic incomes to compute k values of T_2. The sample variance of these k values is the value for c given in the text.

A Simple Adaptation of Diamond's OLG Model

Here we modify the basic Diamond OLG growth model (DOLG) to allow for heterogeneous human capital, which we think of as the amount of accrued time spent in full-time education. We will show that under a null that agents have equal chance of termination of this accumulation process the Theil statistic is approximately a test of fairness in this economy.

We briefly revise the standard textbook model first. (See, for example, Romer, 2001, for a fuller description.) In the standard DOLG agents live for two 'periods' so at any one time we only have the 'young' and 'old'. Young people obtain a certain amount of (randomly drawn years of) education at a cost c_t per year and then supply a fixed amount of labour normalised to unity per individual. Hence Diamond's first period may be thought of as consisting of an early (education) period and a later work period. There are two caveats in the modelling here if we wish to map our exponential (stopping rule) for educational attainment into real time. First, those educated longer would have less time when young to work. Second, the exponential has (positive) infinite support, whereas in reality the amount of time spent in education is finite. With regard to the second factor, we would simply argue that time in education is approximately exponential (truncated at some maximum value). And, for the second, we may specify that a fraction of each unit of education is 'burned' before it is brought to the workplace in order to approximate the opportunity cost of time spent accruing it. As we noted above in the text where we allow for education costs, multiplying total accrued education by a ('burn') factor less than unity does not change the pdf of income. They must invest in one-period lived capital to finance consumption when old (retired). There are no bequests—an assumption that sits well with our desire to build a fair (null) economy.

The model allows for possible population growth and total factor productivity (TFP) growth. Production is achieved via a standard TFP augmented CRS production function and utility is of the usual CRRA form;

$$Y_t = F(K_t, A_t L_t),$$

$$U_{it} = \frac{C_{yt}^{1-\gamma}}{1-\gamma} + \frac{1}{1-\rho}\frac{C_{ot+1}^{1-\gamma}}{1-\gamma},$$

where $C_{yt}^{1-\gamma}(C_{ot}^{1-\gamma})$ is the consumption of each person when young (old), K_t is aggregate capital stock, L_t is the number of young in t and A_t is *TFP* so that $A_t L_t$ is the number of effective units of labour. In this model total income of the young Y_{yt} is just the wage per efficiency unit at t, w_t times the number of efficiency units of labour, A_t

$$Y_{yt} = A_t w_t.$$

Savings of the young and hence their total income when retired is proportional to current (wage) income

$$Y_{ot+1} = s(r_{t+1})Y_{iyt}.$$

In the steady state the levels of capital and income per unit of effective labour are constant. Under log utility ($\gamma = 1$) and a Cobb Douglas production function the steady state is globally stable.

Now we extend the model to allow workers to be heterogeneous in terms of the effective units of labour they supply. To put this another way, we allow individuals to draw productivities a_{it} from an exponential distribution and define A_t in the original model as the sum of these productivities. Differences in a_{it} across individuals arise from individuals having different values of the amount of education they received. The income of individual i when he/she is young (y_{iyt}) and old (y_{iot}) are respectively

$$y_{iyt} = a_{it} w_t \quad => Y_{yt} = \sum_{i=1}^{n} a_{it} w_t = A_t w_t,$$

$$y_{iot+1} = s(r_{t+1}) y_{iyt} \quad => Y_{ot+1} = \sum_{i=1}^{n} y_{iot+1} = s(r_{t+1}) \sum_{i=1}^{n} Y_{iyt} = s(r_{t+1}) Y_{yt}.$$

In effect, if individual i has any extra/less education than individual j the number of his/her efficiency units is merely scaled up/down while the macroeconomic properties of the original model are preserved. This extension allows for a general level of wage premium for education and for cost of education. In particular, if we assume that education costs c_{ht} per unit[28] and that the human capital premium in the wage is $p_{ht} > c_{ht}$ per unit of education h_t acquired then

$$a_{it} = (p_{ht} - c_{ht}) h_{it}.$$

Under the stopping process for education and the random allocation of a fixed ration of education units process described above, h_{it} will be exponentially and multinomially distributed respectively in a cross-section of the young at t. Both a_{it} and y_{iyt} are proportional to h_t and

[28] The cost could represent a payment to an outside educational body made during the first period by the young in exchange for education services. Note that the factor c_{ht} may also allow the crude modelling of the oportunity cost of extra time spent in education.

follow the same distribution. Note that at time t both the income of the old and young have the same distribution as h_t. In an empirical exercise that computes the test for fairness we could use incomes of workers in any age group in any year, but we should obviously not mix incomes from different age groups.

Dyson School and Economics Department, Cornell University
School of Economics, University of Edinburgh

References

Atkinson, A.B. (1970). 'On the measurement of inequality', *Journal of Economic Theory*, vol. 2(3), pp. 244–63.
Blundell, R. (2014). 'Income dynamics and life-cycle inequality: mechanisms and controversies', ECONOMIC JOURNAL, vol. 124(May), pp. 289–318.
Ben-Porath, E., Gilboa, I. and Schmeidler, D. (1997). 'On the measurement of inequality under uncertainty', *Journal of Economic Theory*, vol. 75(1), pp. 194–204.
Bourguignon, F., Ferreira, F.H.G. and Walton, M. (2006). 'Equity, efficiency and inequality traps: a research agenda', *Journal of Economic Inequality*, vol. 5(2), pp. 235–56.
Broome, J. (1984). 'Uncertainty and fairness', ECONOMIC JOURNAL, vol. 94(375), pp. 624–32.
Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014). 'Where is the land of opportunity? The geography of intergenerational mobility in the United States', *Quarterly Journal of Economics*, vol. 129(4), pp. 1553–623.
Cowell, F.A. (2011). *Measuring Inequality*, Oxford: Oxford University Press.
Corak, M. (2013). 'Income inequality, equality of opportunity, and intergenerational mobility', *Journal of Economic Perspectives*, vol. 27(3) pp. 79–102.
De Barros, R. Ferreira, F.H.G. Molinas Vega, J.R. and Chanduvi, J.S. (2009). *Measuring Inequality of Opportunities in Latin America and the Caribbean*, Washington, D.C.: The World Bank.
Diamond, P. (1967). 'Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: comment', *Journal of Political Economy*, vol. 75(5), pp. 765.
Dworkin, R. (1981). 'What is equality? Part 2: Equality of Resources', *Philosophy & Public Affairs*, vol. 10(4), pp. 283–345.
Ferreira, F.H.G. and Gignoux, J. (2011). 'The measurement of inequality of opportunity: theory and an application to Latin America', *Review of Income and Wealth*, vol. 57(4), pp. 622–57.
Fleurbaey, M. and Maniquet, F. (2012). *Equality of Opportunity: The Economics of Responsibility*, World Scientific.
Fleurbaey, M., (2018). 'Welfare economics, risk and uncertainty', *Canadian Journal of Economics*, vol. 51(1), pp. 5–40.
Friedman, M. (1962). *Capitalism and Freedom*, University of Chicago Press.
Huff Stevens, A. (1997). 'Persistent effects of job displacement: the importance of multiple job losses', *Journal of Labor Economics*, vol. 15(1 part1), pp. 165–88.
Kanbur, R. (1987). 'The standard of living: uncertainty, inequality and opportunity', in (G. Hawthorn, ed.), *Amartya Sen's Tanner Lectures: The Standard of Living*, pp. 59–69, Cambridge University Press.
Kanbur, R. (2015). 'Inequality of opportunity: the new motherhood and apple pie? (with A. Wagstaff)', *Health Economics*, vol. 24(10), pp. 1243–7.
Kanbur, R. and Stiglitz, J.E. (2016). 'Dynastic inequality, mobility and equality of opportunity', *Journal of Economic Inequality*, vol. 14(4), pp. 419–34.
Kanbur, R. and Wagstaff, A. (2016). 'How useful is inequality of opportunity as a policy construct?', in (J.E. Stiglitz and K. Basu, eds.), *Inequality and Growth: Patterns and Policy: Volume I: Concepts and Analysis*, pp. 131–50, London, UK: Palgrave Macmillan.
Kass, R.E. and Raftery, A. (1995). 'Bayes factors', *Journal of the American Statistical Association*, vol. 90(430), pp. 773–95.
Lancaster, T. (1979). 'Econometric methods for the duration of unemployment', *Econometrica*, vol. 47(4), pp. 939–56.
Lefranc, A. and Trannoy, A. (2017). 'Equality of opportunity, moral hazard and the timing of luck', *Social Choice and Welfare*, vol. 49(3-4), pp. 469–97.
Mariotti, M. and Veneziani, R. (2017). 'Opportunities as chances: maximising the probability that everybody succeeds', ECONOMIC JOURNAL, vol. 128(611), pp. 1609–33.
Nickell, S. (1979). 'Estimating the probability of leaving unemployment', *Econometrica*, vol. 47(5), pp. 1249–66.
Niehues, J. and Peichl, A. (2014). 'Lower and upper bounds of unfair inequality: theory and evidence for Germany and the US', *Social Choice and Welfare*, vol. 43, pp. 73–99.
Roemer, J. (1998). *Equality of Opportunity*, Cambridge, MA: Harvard University Press.
Roemer, J.E. and Trannoy, A. (2015). 'Equality of opportunity', in (A.B. Atkinson and F. Bourguignon, eds.), *Handbook of Income Distribution, vol. 2*, pp. 217–300, Amsterdam: Elsevier.
Romer, D. (2001). *Advanced Macroeconomics*, McGraw-Hill.

Saito, K. (2013). ‘Social preferences under risk: equality of opportunity versus equality of outcome’, *American Economic Review*, vol. 103(7), pp. 3084–101.

Sen, A. (1980). *Equality of What? The Tanner Lectures on Human Values*, vol. 1, Cambridge University Press.

Shannon, C.E. (1948). ‘A mathematical theory of communication’, *Bell System Technical Journal*, vol. 27(3), pp. 379–423.

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