

# Inequality, Poverty, and Stochastic Dominance

by

**Russell Davidson**

Department of Economics and CIREQ  
McGill University  
Montréal, Québec, Canada  
H3A 2T7

AMSE and GREQAM  
Centre de la Vieille Charité  
2 rue de la Charité  
13236 Marseille cedex 02, France

email: [russell.davidson@mcgill.ca](mailto:russell.davidson@mcgill.ca)

February 2017

## The First Step: Corrado Gini

Corrado Gini (May 23, 1884 – 13 March 13, 1965) was an Italian statistician, demographer and sociologist who developed the Gini coefficient, a measure of the income inequality in a society. Gini was born on May 23, 1884 in Motta di Livenza, near Treviso, into an old landed family. He entered the Faculty of Law at the University of Bologna, where in addition to law he studied mathematics, economics, and biology. In 1929 Gini founded the Italian Committee for the Study of Population Problems (*Comitato italiano per lo studio dei problemi della popolazione*) which, two years later, organised the first Population Congress in Rome. In 1926 he was appointed President of the Central Institute of Statistics in Rome. This he organised as a single centre for Italian statistical services. He resigned in 1932 in protest at interference in his work by the fascist state.

## The relative mean absolute difference

Corrado Gini himself (1912) thought of his index as the average difference between two incomes, divided by twice the average income. The numerator is therefore the double sum

$$\sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|,$$

where the  $y_i$ ,  $i = 1, \dots, n$  are the incomes of a finite population of  $n$  individuals, divided by the number of income pairs. There has subsequently been disagreement, sometimes acrimonious, in the literature as to the appropriate definition of the number of pairs. Should a single income constitute a pair with itself or not? If so, the number of pairs is  $n^2$ ; if not,  $n(n - 1)$ .

It turns out that, for some purposes, one definition is better, but for other purposes, it is the other definition that is better.

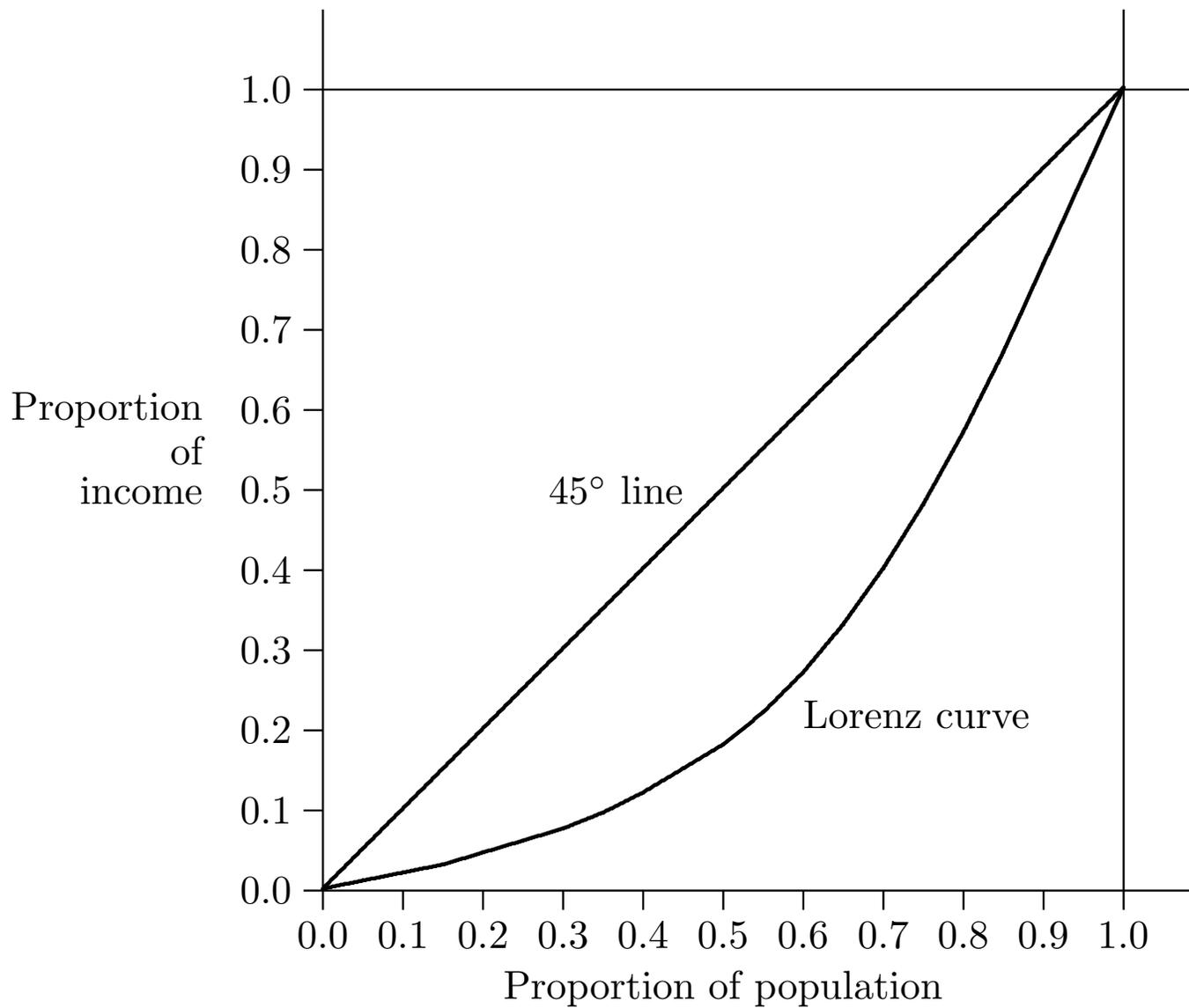
The average income is of course just  $\mu \equiv n^{-1} \sum_{i=1}^n y_i$ , and so one definition of the index is

$$G = \frac{1}{2\mu} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|,$$

## Lorenz Curves and Pen's Parade

The traditional way of displaying the size distribution graphically is by a Lorenz curve. To understand how it is constructed, think of something called Pen's income parade. Everyone in the population is lined up in order of income, with the poorest people at the head of the parade, and the richest at the end. Pen, a Dutch economist writing in the middle of the last century, warned us that, strictly speaking, the parade is headed by people with huge negative incomes, those who lost a pile of money on the stock exchange. Let's forget about that.

As the parade goes past, we record the cumulative sum of the incomes of the people who have gone by. Once ten per cent, say, have passed us, we have a count of how much income accrues to the ten percent poorest people. We carry on this way until we have counted everyone's income and totalled it. We can then divide the cumulative sums at 10, 20, 30, per cent, and so on, by total income to get income shares. The Lorenz curves plots these shares on the  $y$  axis and the percent of the population on the  $x$  axis. Because of how it is constructed, the Lorenz curve always lies below the 45 degree line in the plot, unless income is absolutely equally distributed, when it coincides with the 45 degree line. The curve lets us see at a glance how unequal the distribution is. If it lies close to the 45 degree line, income is distributed rather equally, if not, inequality is significant.



**A Typical Lorenz Curve**

The definition that I learned as a student is that the Gini index is twice the area between the 45°-line and the Lorenz curve. The Lorenz curve itself is defined implicitly by

$$L(F(x)) = \frac{1}{\mu} \int_0^x y \, dF(y), \quad x \in [0, \infty),$$

where  $F$  is the population CDF. Thus

$$\begin{aligned} G &= 1 - 2 \int_0^1 L(y) \, dy = 1 - 2 \int_0^\infty L(F(x)) \, dF(x) \\ &= 1 - \frac{2}{\mu} \int_0^\infty dF(x) \int_0^x y \, dF(y) \\ &= 1 - \frac{2}{\mu} \left[ F(x) \int_0^x y \, dF(y) \right]_0^\infty + \frac{2}{\mu} \int_0^\infty F(x)x \, dF(x) \\ &= 1 - 2 + \frac{2}{\mu} \int_0^\infty xF(x) \, dF(x) = \frac{2}{\mu} \int_0^\infty xF(x) \, dF(x) - 1 \end{aligned}$$

There are many other equivalent expressions, but this one is as convenient as any. In principle, it can be used for either a continuous population or a finite, discrete, one.

## Equivalence of the Definitions

At first glance, there is little connection between the definition of the sample Gini as just defined and the definition in terms of the mean absolute deviation. For a sample with **empirical distribution function (EDF)**  $\hat{F}$ , the sample Gini as just defined is

$$\frac{2}{\hat{\mu}} \int_0^{\infty} y \hat{F}(y) d\hat{F}(y) - 1,$$

where  $\hat{\mu}$  is the sample mean. We can express the mean absolute deviation as follows:

$$\begin{aligned} & \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |y_j - y_i| = \frac{2}{n^2} \sum_{y_i < y_j} (y_j - y_i) \\ &= \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n (y_j - y_i) \mathbf{I}(y_i < y_j) \\ &= \frac{2}{n} \left[ \sum_{j=1}^n y_j \hat{F}(y_j) - \sum_{i=1}^n y_i (1 - \hat{F}(y_i)) \right] \\ &= 4 \int_0^{\infty} y \hat{F}(y) d\hat{F}(y) - 2\hat{\mu} \end{aligned}$$

Divide by  $2\hat{\mu}$  and the equivalence follows.

## Definitional Issues

An EDF  $\hat{F}$  is necessarily discontinuous, and it gives a different value for the formula we have just derived according as  $\hat{F}$  is defined as left- or right-continuous. That is, is  $\hat{F}$  cadlag or is it caglad (ladcag)? The usual, purely conventional, choice is that it is cadlag. The two different answers are

$$\frac{1}{n^2\mu} \sum_{i=1}^n 2iy_{(i)} - 1 \quad \text{or} \quad \frac{1}{n^2\mu} \sum_{i=1}^n (2i - 2)y_{(i)} - 1.$$

Here the  $y_{(i)}$ ,  $i = 1, \dots, n$ , are the **order statistics** of the sample of which  $\hat{F}$  is the EDF. If we evaluate the earlier definition in terms of the mean absolute deviation, with a denominator of  $n^2$ , the result is the average of these. Using this average has two advantages over all other potential definitions:

- If there is perfect income equality,  $y_i = \mu$  for all  $i$ , then  $G = 0$ .
- It satisfies the **population symmetry axiom**, whereby, if the population is exactly replicated, the value of  $G$  does not change.

It therefore seems best to adopt this definition for finite populations, and also for the **sample Gini** for a sample drawn from a population that can be either continuous or discrete.

## Sen's Axioms

Amartya Sen (1976) focussed more on poverty than on inequality. In discussing poverty, a necessary concept is the **poverty line**, an income level such that individuals whose income is lower than this line are deemed to be poor.

A very straightforward, and frequently used, poverty measure is the **headcount ratio**, which is simply the number of poor people divided by the size of the entire population. We have  $H = F(z)$ , where  $F$  is the CDF of the income distribution,  $z$  is the poverty line, and  $H$  is the headcount ratio. Sen considered the headcount ratio a very bad measure, since it takes no account of the **intensity** of poverty.

Sen therefore enunciated some axioms that he thought a good measure of poverty should satisfy.

- *Monotonicity*

Given other things, a reduction in income of a person below the poverty line must increase the poverty measure.

- *Transfers*

Given other things, a pure transfer from a person below the poverty line to someone who is richer must increase the poverty measure.

For an income  $y$ , the **income gap**, or **poverty gap**, is defined as  $g \equiv z - y$ , the amount by which the income  $y$  must be augmented to bring it up to the poverty line. Sen begins by defining a very general class of aggregate poverty measures. Let  $x$  be an arbitrary income level. The aggregate gap of all people with income less than  $x$ , when the poverty line is  $z$ , is

$$Q(x) = A(z, \mathbf{y}) \sum_{y_i < x} g_i v_i(z, \mathbf{y}).$$

Here  $\mathbf{y}$  is the vector of all incomes in the population. It can be seen that  $Q(x)$  is a weighted average of the poverty gaps of people with income less than  $x$ . The overall coefficient  $A(z, \mathbf{y})$ , and the weights  $v_i(z, \mathbf{y})$ , are so far quite general, except for being required to be non-negative, but they will be progressively restricted by the axioms that follow.

Given the definition of  $Q$ , then the corresponding poverty index is  $P = Q(z)$ . It is necessary for the statement of some of the axioms to suppose that we have some sort of welfare measure for people, in general dependent on the vector  $\mathbf{y}$  of all incomes. Let the welfare of individual  $i$  be denoted as  $W_i(\mathbf{y})$ .

- *Relative Equity*

For any pair  $i, j$ : if  $W_i(\mathbf{y}) < W_j(\mathbf{y})$ , then  $v_i(z, \mathbf{y}) > v_j(z, \mathbf{y})$ .

We do *not* require  $W$  to be a **cardinal** measure. All that we need is to be able to tell that  $i$  is worse off than  $j$ . However, the next axiom, which implies the relative equity axiom, imposes more.

- *Ordinal Rank Weights*

The weight  $v_i(z, \mathbf{y})$  is equal to the **rank-order** of  $i$  in the interpersonal welfare ordering of the poor.

Among other things, we require for this that we have a **complete** welfare ordering, at least of the poor. The next axiom extends this requirement, and links welfare to income.

- *Monotonic Welfare*

The relation  $>$  defined on the set of individual welfares,  $\{W_i(\mathbf{y})\}$ , for any income configuration  $\mathbf{y}$ , provides a strict complete ordering. If  $y_i < y_j$ , where  $<$  is the usual relation defined on the real numbers, then  $W_i(\mathbf{y}) < W_j(\mathbf{y})$ .

For poverty line  $z$ , let  $q = nF(z)$  be the number of poor individuals. The **income-gap ratio** can be thought of as a per-person proportional gap. It is defined as

$$I = \sum_{y_i < z} g_i / (qz),$$

The headcount ratio is just  $F(z) = q/n$ . It is insensitive to the depth, or intensity, of poverty, but the income-gap ratio is insensitive to the proportion of poor people in the population – note that  $n$  appears nowhere in its definition. A decent poverty measure should depend on both  $H$  and  $I$ , but that is still not enough, *unless* all the poor have exactly the same income. In this special case, if we set  $P = HI$ , the axioms set out so far are satisfied, and we can have the following:

- *Normalised Poverty Value*

If the poor all receive exactly the same income, then  $P = HI$ .

### Theorem (Sen)

For  $q$  large enough, the only poverty index satisfying the Axioms on ordinal rank weights, monotonic welfare, and normalised poverty value is

$$P = H\left(I + (1 - I)G\right),$$

where  $G$  is the Gini index for the poor segment of the population.

### Proof:

Order incomes, and hence also welfares, so that  $i < j$  implies  $y_i < y_j$ . Thus the  $y_i$  are the order statistics of the income distribution. By requiring ordinal rank weights, we see that  $v_i(z, \mathbf{y}) = q - i + 1$ . Then

$$P = A(z, \mathbf{y}) \sum_{i=1}^q g_i(q - i + 1).$$

Note that, letting  $j = q - i + 1$ ,

$$\sum_{i=1}^q (q - i + 1) = \sum_{j=1}^q j = q(q + 1)/2.$$

If all the poor have the same income gap  $g^*$ , then we have

$$P = A(z, \mathbf{y})g^* \sum_{i=1}^q (q - i + 1) = A(z, \mathbf{y})g^* q(q + 1)/2.$$

By the requirement of Normalised Poverty Value, if  $g_i = g^*$  for all poor individuals  $i$ , we must have  $P = HI = (q/n)(g^*/z)$ . Hence

$$A(z, \mathbf{y}) = \frac{2}{nz(q + 1)},$$

and so, in general,

$$\begin{aligned} P &= \frac{2}{nz(q + 1)} \sum_{i=1}^q (z - y_i)(q - i + 1) \\ &= \frac{q}{n} - \frac{2}{nz(q + 1)} \sum_{i=1}^q y_i(q - i + 1) = \frac{q}{n} \left[ 1 - \frac{2}{zq(q + 1)} \sum_{i=1}^q y_i(q - i + 1) \right]. \end{aligned}$$

Now recall that  $q/n = H$ , and

$$I = \sum_{i=1}^q g_i / (qz) = \frac{1}{qz} \sum_{i=1}^q (z - y_i) = 1 - \mu_p / z,$$

where  $\mu_p$  is the average income of the poor. Further, using one of the expressions for the Gini index  $G$  for the poor only, we see that

$$G = \frac{1}{q^2 \mu_p} \sum_{i=1}^q (2i - 2)y_i - 1,$$

from which it follows that

$$2 \sum_{i=1}^q y_i (i - 1) = q^2 \mu_p (G + 1).$$

From these relations, we see that

$$\begin{aligned}
 P &= \frac{q}{n} \left[ 1 - \frac{2}{zq(q+1)} \sum_{i=1}^q y_i (q - i + 1) \right] \\
 &= H \left[ 1 - \frac{2}{z(q+1)} q\mu_p + \frac{q}{z(q+1)} \mu_p (G + 1) \right] = H \left[ 1 + \frac{q\mu_p}{z(q+1)} (G - 1) \right] \\
 &= H \left[ 1 + \frac{q}{q+1} (1 - I)(G - 1) \right].
 \end{aligned}$$

Now, as  $q \rightarrow \infty$ ,  $q/(q+1) \rightarrow 1$ , and so for large enough  $q$ , this expression is approximately

$$H[1 + (1 - I)(G - 1)] = H[1 + G - 1 - IG + I] = H[I + (1 - I)G].$$



## Blackorby and Donaldson: Ethical Content

It may be clear from the reasoning of the previous section that the axiom that leads to a *unique* poverty index is the ordinal rank weights axiom. This axiom is uncomfortably strong. It corresponds to a very particular weighting of the incomes of the poor. The axiom can of course be relaxed, but then we no longer have any criterion to judge among the different indices that satisfy the other axioms.

If one is prepared to introduce a **social welfare function** in order to have a criterion for making ethical judgements of the sort needed, then this introduces the possibility of interpersonal comparisons stronger than those needed to say only whether  $i$  is better or worse off than  $j$ . It allows us to say whether income distribution  $A$  is better than income distribution  $B$ , and it allows us to define the **representative income** of any subgroup in society, for instance, the poor as defined by some poverty level  $z$ . If we denote the social welfare function (SWF) by  $W$ , then the representative income of a particular subgroup  $S$  of the population is that income which, if given equally to all members of  $S$ , leaves the value of the SWF unchanged from its value for the original vector  $\mathbf{y}$  of incomes.

- *Anonymity*

Social welfare is unaffected by permuting the incomes of the individuals in the society.

This axiom is obviously related to some of the ideals of modern society. It would certainly not be respected by policy makers in the Middle Ages, or indeed by any modern government that respected class or caste distinctions. The formal consequence of the axiom, which thus allows us to permute the elements of the income vector  $\mathbf{y}$ , is that social welfare can be considered as a function only of the order statistics defined by  $\mathbf{y}$ . In what follows, we assume that the elements of  $\mathbf{y}$  are ordered from smallest to greatest.

If we take all the worldly goods of a rich man, call him Bill Gates, or Dives if you prefer biblical allusions, and give them to a poor man, such as a graduate student, or the biblical Lazarus, and give the few worldly goods of Lazarus to Dives, then we should give equal rankings to the situations before and after. As the participants in Pen's parade do not call out their names as they pass, it is clear that anything based on Lorenz curves, including the Gini coefficient, respect the requirement of anonymity.

Consider the social subgroup of the poor as defined by a poverty line  $z$ , and let  $\mathbf{y}$  be truncated by keeping only the incomes less than  $z$ . The welfare of the poor is measured by  $W(\mathbf{y})$ , and their representative income  $\xi$  is defined by

$$W(\mathbf{y}) = W(\xi\boldsymbol{\iota}),$$

where  $\boldsymbol{\iota}$  is the vector every element of which is 1.

Blackorby and Donaldson (BD) define an index of inequality of the poor as

$$I(\mathbf{y}) = (\mu_p(\mathbf{y}) - \xi) / \mu_p(\mathbf{y}).$$

This index is positive if  $W$  is concave (and also under weaker conditions), because, if the total income of the poor is divided up equally among them, so that each receives the mean income  $\mu_p(\mathbf{y})$ , this configuration has higher welfare because it has less inequality. Therefore the representative income in the unequal configuration is lower.

Clearly the functional form of  $I$  can be computed knowing  $W$ , and this holds also in the other direction. To see this, it is convenient to make the assumption that  $W$  is homogeneous of degree one. This is not very restrictive, since  $W$  is thought of as an **ordinal** index. It means that we may start from any homothetic SWF, and use a homogeneous function that is ordinally equivalent to it. In addition, we may suppose without loss of generality that  $W(\boldsymbol{\iota}) = 1$ .

Then, if  $W$  is homogeneous,  $W(\xi\boldsymbol{\iota}) = \xi W(\boldsymbol{\iota}) = \xi$ . Since  $\xi = \mu_p(\mathbf{y})(1 - I(\mathbf{y}))$ , we see that

$$W(\mathbf{y}) = W(\xi\boldsymbol{\iota}) = \xi = \mu_p(\mathbf{y})(1 - I(\mathbf{y})).$$

## Sen revisited

Suppose that the inequality index  $I(\mathbf{y})$  is the Gini index of the distribution of the incomes of the poor. In order to avoid notational confusion, we will denote the income-gap ratio by  $J$  rather than  $I$ . Then  $J = 1 - \mu_p(\mathbf{y})/z$ . Sen's index is

$$\begin{aligned} H(J + (1 - J)G) &= H\left[1 - \frac{\mu_p(\mathbf{y})}{z} + \frac{\mu_p(\mathbf{y})}{z}I(\mathbf{y})\right] \\ &= H\left[1 - \frac{\mu_p(\mathbf{y})}{z}(1 - I(\mathbf{y}))\right]. \end{aligned}$$

But  $I(\mathbf{y}) = 1 - \xi/\mu_p(\mathbf{y})$ , and so Sen's index is

$$H\left[1 - \frac{\xi}{z}\right] = H\left[\frac{z - \xi}{z}\right].$$

## Sen generalised

The functional form assumed by Sen's index is very neat. If the representative income  $\xi$  is defined using a SWF different from the one implicitly defined by the Gini index, we can still use the same functional form for an inequality index. Although BD begin by suggesting a still more general index of the form

$$f\left(H, \frac{z - \xi}{z}\right),$$

for some so-far unspecified function  $f$ , they note that, if  $f(H, I) = HI$ , then doubling the headcount ratio doubles the index, and, similarly, doubling the proportional income gap  $(z - \xi)/z$  also doubles the index. These properties make much intuitive sense.

## Axioms for Ethics

BD argued that the use of the Gini coefficient, or any of the possible generalised Gini coefficients, was *not* suitable for the measurement of either poverty or inequality, since the SWF on which it implicitly relies does not satisfy some of the ethically determined requirements for a suitable SWF. The first of these we have already discussed, and it is not a problem – it is simply the **anonymity** of individuals.

Another requirement is the **Pareto principle**.

- *Pareto Principle*

Situation  $B$  is better than situation  $A$  if at least one individual is better off in  $B$  than in  $A$ , and no one is worse off.

Use of a Lorenz curve does *not* in general respect the Pareto principle. If we increase the income of the richest member of society, then Pareto tells us that society's welfare has increased. Inequality, however, has increased as well, and so the Gini coefficient has become larger. This illustrates a feature of the Lorenz curve, which is that it is invariant not only under changes in the units of measurement of income, but also under proportional changes in people's real incomes. It thus distinguishes inequality from welfare, in the sense that, provided everyone has the same income, the Lorenz curve has reached its bliss point, even if everyone is in abject poverty. In order for a SWF to respect the Pareto principle, it must be increasing in all of its arguments.

BD claim that, for a SWF defined on the set of the poor to have what they call ethical content, it must be a specialisation to the poor of a SWF defined on the entire population. If we continue to denote by  $\mathbf{y}$  the vector of poor incomes, and let  $\mathbf{Y}$  be the vector of all the others, then a SWF defined on the population can be denoted as  $\mathcal{W}(\mathbf{y}, \mathbf{Y})$ . If  $\mathbf{Y}$  is fixed at  $\bar{\mathbf{Y}}$ , then we can conclude that what we want is  $\mathcal{W}(\mathbf{y}) = \mathcal{W}(\mathbf{y}, \bar{\mathbf{Y}})$ .

- *Independence from the non-Poor*

A poverty index is invariant under changes in the distribution of the non-poor.

The definition above does not in general satisfy this axiom, as it appears that a change in  $\mathbf{Y}$  can affect, not only the unimportant numerical value of the welfare of the poor, but also the ordering of different values of  $\mathbf{y}$ . To avoid this, we require that the arguments  $\mathbf{Y}$  of the function  $\mathcal{W}$  should be **separable** from  $\mathbf{y}$ , which means that there are functions  $\bar{\mathcal{W}}$  and  $W_1$  such that

$$\mathcal{W}(\mathbf{y}, \mathbf{Y}) = \bar{\mathcal{W}}(\mathbf{y}, W_1(\mathbf{Y})).$$

Even this is specific to a particular choice of the poverty line  $z$ . If we wish to maintain separability for arbitrary  $z$ , then a sufficient condition for this is to have the function  $\mathcal{W}$  be **additively separable** in its arguments. This means that

$$\mathcal{W}(y_1, \dots, y_n) = \sum_{i=1}^n u_i(y_i).$$

In addition, anonymity requires that there is just one function  $u$ :  $u_i = u$  for all  $i$ .

*Stochastic Dominance*

## Statistical Inference for Lorenz Curves

In essentially all work on statistical inference in this course, the central idea is to express the quantity of interest as a sum of IID random variables, at least asymptotically. We can then use a law of large numbers to investigate consistency, and a central-limit theorem to investigate asymptotic normality and find expressions for asymptotic covariance matrices. In favourable cases, such an asymptotic covariance matrix can be estimated consistently without making parametric assumptions.

Let us begin with a simple example. If we have an IID sample  $\{y_i\}$ ,  $i = 1, \dots, n$ , the EDF of the sample,  $\hat{F}$  say, estimates the underlying distribution consistently. The definition of  $\hat{F}$  is

$$\hat{F}(x) = n^{-1} \sum_{i=1}^n \mathbf{I}(y_i \leq x),$$

with  $\mathbf{I}$  the indicator function. Here the expression on the right-hand side is an average of IID realisations from the distribution of the random variable  $\mathbf{I}(Y \leq x)$ , where  $Y \sim F$ ,  $F$  being the CDF of the distribution. Note first that

$$\mathbf{E}(\mathbf{I}(Y \leq x)) = P(Y \leq x) = F(x),$$

so that, by the LLN,  $\hat{F}(x) \rightarrow F(x)$  in probability, and indeed almost surely. The result holds for all  $x$  where  $F$  is continuous, but fails to hold at points of discontinuity on account of the fact that convention requires a CDF to be cadlag.

For asymptotic normality, we look at the estimation error  $\hat{F}(x) - F(x)$ , scaled up by a factor of  $n^{1/2}$ . We see that

$$n^{1/2} \sum_{i=1}^n (\hat{F}(x) - F(x)) = n^{-1/2} \sum_{i=1}^n (\mathbf{I}(y_i \leq x) - F(x)).$$

The expectation of this is zero, and the variance is

$$\text{Var}(\mathbf{I}(Y \leq x) - F(x)) = \mathbf{E}(\mathbf{I}(Y \leq x)) - F^2(x) = F(x)(1 - F(x)).$$

Here we used the fact that the square of the indicator function is just the indicator function itself – it is idempotent. The CLT tells us that  $n^{1/2}(\hat{F}(x) - F(x))$  is asymptotically normal with expectation zero and variance  $F(x)(1 - F(x))$ .

But more generally, we are interested in the asymptotic covariance of  $\hat{F}(x_1)$  and  $\hat{F}(x_2)$ , with  $x_1 \leq x_2$ . By definition, the covariance is

$$n^{-1} \mathbf{E}[(\mathbf{I}(Y \leq x_1) - F(x_1))(\mathbf{I}(Y \leq x_2) - F(x_2))],$$

which is  $n^{-1}$  times

$$\begin{aligned} & \mathbf{E}(\mathbf{I}(Y \leq x_1)) - F(x_1)\mathbf{E}(\mathbf{I}(Y \leq x_2)) - F(x_2)\mathbf{E}(\mathbf{I}(Y \leq x_1)) + F(x_1)F(x_2) \\ & = F(x_1)(1 - F(x_2)). \end{aligned}$$

## A Lorenz Ordinate

We defined the Lorenz curve as the graph of the function defined implicitly by

$$L(F(x)) = \frac{1}{\mu} \int_0^x y \, dF(y).$$

If we let  $p = F(x)$ , then the ordinate of the Lorenz curve for abscissa  $p$  is

$$L(p) = \frac{1}{\mu} \int_0^{Q(p)} y \, dF(y),$$

where  $Q$  is the **quantile function**. We assume here that  $F$  is continuous and strictly increasing on its support, which means that  $Q$  is uniquely defined and is such that

$$F(Q(p)) = p \quad \text{and} \quad Q(F(x)) = x.$$

Since  $Y$  is income, we suppose the support of  $Y$  is a subset of the non-negative real line. Define  $\gamma(p)$  to be such that  $p\gamma(p) = \mathbb{E}(Y \mathbb{I}(Y \leq Q(p)))$ . Then  $\gamma(1) = \mathbb{E}(Y) = \mu$ . We now want to estimate  $\gamma(p)$  for given  $p$ , and to compute the asymptotic covariance of  $\gamma(p_1)$  and  $\gamma(p_2)$  with  $p_1 \leq p_2$ .

The definition of  $\gamma$  can be written as

$$p\gamma(p) = \int_0^\infty y \mathbf{I}(y \leq Q(p)) \, dF(y) = \int_0^{Q(p)} y \, dF(y).$$

The **plug-in** estimator replaces the unknown  $F$  by the EDF  $\hat{F}$ , and the quantile function  $Q$  by  $\hat{Q}$ , an inverse of  $\hat{F}$ . Thus

$$p\hat{\gamma}(p) = \int_0^\infty y \mathbf{I}(y \leq \hat{Q}(p)) \, d\hat{F}(y) = \int_0^{\hat{Q}(p)} y \, d\hat{F}(y).$$

The integral w.r.t.  $\hat{F}$  is in fact a sum. We have

$$p\hat{\gamma}(p) = n^{-1} \sum_{i=1}^n y_i \mathbf{I}(y_i \leq \hat{Q}(p)).$$

In order to deal with the estimated quantile, we proceed as follows.

$$\begin{aligned}
p\hat{\gamma}(p) &= \int_0^{\hat{Q}(p)} y \, d\hat{F}(y) = \int_0^{Q(p)} y \, d\hat{F}(y) + \int_{Q(p)}^{\hat{Q}(p)} y \, d\hat{F}(y) \\
&= n^{-1} \sum_{i=1}^n y_i \mathbf{I}(y_i \leq Q(p)) + \int_{Q(p)}^{\hat{Q}(p)} y \, d\hat{F}(y)
\end{aligned}$$

At this point, we need some asymptotic approximations. We know that  $\hat{F}(y) - F(y) = O_p(n^{-1/2})$ , whence, under our assumptions,  $\hat{Q}(p) - Q(p) = O_p(n^{-1/2})$  as well, and so also the integral from  $Q(p)$  to  $\hat{Q}(p)$ . Then:

$$\begin{aligned}
\int_{Q(p)}^{\hat{Q}(p)} y \, d\hat{F}(y) &= -Q(p) (\hat{F}(Q(p)) - p) + O(n^{-1}) \\
&= pQ(p) - n^{-1}Q(p) \sum_{i=1}^n \mathbf{I}(y_i \leq Q(p)) + O(n^{-1}).
\end{aligned}$$

It follows that, ignoring terms of order  $n^{-1}$ ,

$$p\hat{\gamma}(p) = pQ(p) + n^{-1} \sum_{i=1}^n (y_i - Q(p))I(y_i \leq Q(p)).$$

The expectation of the r.h.s. is

$$pQ(p) + E[(Y - Q(p))I(Y \leq Q(p))] = pQ(p) + p\gamma(p) - pQ(p) = p\gamma(p).$$

We used the fact that  $E[I(Y \leq Q(p))] = P(Y \leq Q(p)) = p$ . Thus  $p\hat{\gamma}(p)$  is asymptotically unbiased and consistent. Now we can see that, to order unity,

$$\begin{aligned} n^{1/2}(p\hat{\gamma}(p) - p\gamma(p)) \\ = n^{-1/2} \sum_{i=1}^n \left[ (y_i - Q(p))I(y_i \leq Q(p)) - E[(Y - Q(p))I(Y \leq Q(p))] \right]. \end{aligned}$$

The covariance of  $n^{1/2}(p_1\hat{\gamma}(p_1) - p_1\gamma(p_1))$  and  $n^{1/2}(p_2\hat{\gamma}(p_2) - p_2\gamma(p_2))$  can now be written down to order unity. Under the assumption that  $p_1 \leq p_2$ , it is

$$\begin{aligned} & \mathbf{E}\left[(Y - Q(p_1))\mathbf{I}(Y \leq Q(p_1))(Y - Q(p_2))\mathbf{I}(Y \leq Q(p_2))\right] \\ & \quad - \mathbf{E}[(Y - Q(p_1))\mathbf{I}(Y \leq Q(p_1))]\mathbf{E}[(Y - Q(p_2))\mathbf{I}(Y \leq Q(p_2))] \\ = & \mathbf{E}\left[(Y - Q(p_1))(Y - Q(p_2))\mathbf{I}(Y \leq Q(p_1))\right] - p_1(\gamma(p_1) - Q(p_1))p_2(\gamma(p_2) - Q(p_2)) \\ = & \mathbf{E}[Y^2\mathbf{I}(Y \leq Q(p_1))] - p_1\gamma(p_1)Q(p_1) - p_1\gamma(p_1)Q(p_2) + p_1Q(p_1)Q(p_2) \\ & \quad - p_1p_2(\gamma(p_1) - Q(p_1))(\gamma(p_2) - Q(p_2)) \end{aligned}$$

If we define  $\lambda^2(p)$  so that  $p(\lambda^2(p) + \gamma^2(p)) = \mathbf{E}[Y^2\mathbf{I}(Y \leq Q(p))]$ , then  $\lambda^2(p)$  is the variance of  $Y$  conditional on  $Y \leq Q(p)$ . After a lot of tidying, the covariance can be written as

$$p_1\lambda^2(p_1) + p_1(1-p_2)(Q(p_1) - \gamma(p_1))(Q(p_2) - \gamma(p_2)) + p_1(Q(p_1) - \gamma(p_1))(\gamma(p_2) - \gamma(p_1)).$$

This formula was derived by Beach and Davidson (1983). Everything in it can be estimated in a **distribution-free** manner, using  $\hat{\gamma}(p_i)$  and  $\hat{Q}(p_i)$ ,  $i = 1, 2$  to estimate  $\gamma(p_i)$  and  $Q(p_i)$  respectively.

## Stochastic Dominance

In an influential paper, Atkinson (1987) raised a number of issues in poverty measurement that he felt were in need of clarification. He raised three particular points.

- What is the appropriate choice of the poverty line?
- The poverty line once chosen, what is the appropriate poverty index?
- What should be the relation between indices of poverty and indices of inequality?

In fact, Blackorby and Donaldson had already given elements of answers to the second two questions in their 1980 paper, of which Atkinson was presumably unaware at the time he wrote his paper, since he does not cite it there.

One way of finessing the first of Atkinson's questions, concerning the choice of the poverty line, is to enquire for what range of poverty lines a comparison of the poverty in two populations always leads to the same answer. Let us begin by considering the headcount ratio. If we have two populations,  $A$  and  $B$ , characterised by two CDFs,  $F_A$  and  $F_B$ , then, for poverty line  $z$ , the headcount ratio is higher in  $A$  than in  $B$  if and only if  $F_A(z) > F_B(z)$ .

Should it be the case that  $F_A(y) > F_B(y)$  for *all*  $y$ , then population  $B$ , or equivalently the distribution  $F_B$ , is said to **dominate**  $A$ , or  $F_A$ , **stochastically at first order**. If the inequality holds for all  $y$  up to a threshold  $z$ , then we have first-order stochastic dominance up to  $z$ . If the inequality holds for all  $y$  in an interval  $[z_-, z_+]$ , then we have **restricted stochastic dominance** in this interval.

It is easy to see that first-order stochastic dominance of  $A$  by  $B$  means that  $B$  has higher social welfare than  $A$  for all additively separable SWFs that respect anonymity. In fact, such a SWF can be written as a Riemann-Stieltjes integral, as follows:

$$W(y_1, \dots, y_n) = \int_0^\infty u(y) dF(y),$$

The right-hand side can be used as the definition of a SWF for a continuous distribution  $F$ . It can also be written as

$$- \int_0^\infty u(y) d(1 - F(y)),$$

since  $d(1 - F(y)) = -dF(y)$ . Integrating by parts, we can see that the SWF becomes

$$\begin{aligned} & - [u(y) (1 - F(y))]_{y=0}^{y=\infty} + \int_0^\infty (1 - F(y)) u'(y) dy \\ & = \int_0^\infty (1 - F(y)) u'(y) dy \end{aligned}$$

if we assume that  $u(0) = 0$  and that  $u$  is differentiable.

Now consider the two populations  $A$  and  $B$ , and compute the difference between the values of the SWF defined by the function  $u$  for the two distributions. We have

$$W_B - W_A = \int_0^{\infty} u'(y) (F_A(y) - F_B(y)) dy.$$

Since  $u$  is required to be an increasing function, we see that  $u'(y) > 0$  for all  $y$ . If the first-order stochastic dominance condition that  $F_A(y) \geq F_B(y)$  for all  $y$  is satisfied, then the difference is necessarily positive. Thus first-order stochastic dominance of  $A$  by  $B$  is a sufficient condition for all anonymous, additively separable, SWFs to agree that  $B$  has more social welfare than  $A$ . In fact, this dominance is also a necessary condition, since, if it does not hold, we can find some increasing function  $u$  that puts more weight on the range over which dominance does not hold than the range over which it does.

Now let us restrict attention to the welfare of people with incomes less than  $z$ . Instead of a function  $u$  that measure social welfare, we use a function  $\pi$  that measures the disutility of the poverty gap. We can define a class of **poverty indices** as follows:

$$\Pi(z) = \int_0^z \pi(z - y) dF(y),$$

where we take account only of people whose incomes are less than  $z$ . Integration by parts gives

$$\Pi(z) = [\pi(z - y) F(y)]_{y=0}^{y=z} + \int_0^z \pi'(z - y) F(y) dy = \int_0^z \pi'(z - y) F(y) dy,$$

where we assume that  $\pi(0) = 0$ . The difference between the values of this expression for populations  $A$  and  $B$  is

$$\int_0^z \pi'(z - y) (F_A(y) - F_B(y)) dy,$$

which is positive when  $B$  dominates  $A$  at first order, provided that  $\pi' > 0$ , which just means that disutility increases with an increase of the poverty gap. We conclude that, for all poverty indices  $\Pi(z)$ , there is more poverty in  $A$  than in  $B$  if  $B$  dominates  $A$  at first order.

For a given CDF  $F$ , let us define the sequence of functions  $D^s$ ,  $s = 1, 2, \dots$ , by the recurrence relation

$$D^1(z) = F(z), \quad D^{s+1}(z) = \int_0^z D^s(y) \, dy.$$

A distribution  $B$  dominates a distribution  $A$  stochastically at order  $s$  if  $D_A^s(z) \geq D_B^s(z)$  for all  $z$ . It is clear that dominance at any given order  $s$  implies dominance for all higher orders. Restricted dominance at order  $s$  over an interval can be defined exactly as for first-order dominance.

It is often convenient to have an explicit representation of the functions  $D^s$ . This representation is as follows:

$$D^s(z) = \frac{1}{(s-1)!} \int_0^z (z-y)^{s-1} \, dF(y).$$

This is clearly true for  $s = 1$ , if we remember that  $0! = 1$ . Then suppose that it holds for some given  $s$ ; we show that it holds for  $s + 1$ , which allows us to assert the general result by induction. We have from the recurrence relation that

$$D^{s+1}(z) = \int_0^z D^s(y) \, dy = \frac{1}{(s-1)!} \int_0^z dy \int_0^y (y-x)^{s-1} \, dF(x).$$

We interchange the order of integration (Fubini's theorem) and find that

$$\begin{aligned}
 D^{s+1}(z) &= \frac{1}{(s-1)!} \int_0^z dF(x) \int_x^z (y-x)^{s-1} dy \\
 &= \frac{1}{(s-1)!} \int_0^z dF(x) \int_0^{z-x} w^{s-1} dw \\
 &= \frac{1}{s(s-1)!} \int_0^z dF(x) (z-x)^s = \frac{1}{s!} \int_0^z (z-x)^s dF(x),
 \end{aligned}$$

as required.

For  $s = 2$ , we have

$$D^2(z) = \int_0^z (z-y) dF(y),$$

from which we see that, for given  $z$ ,  $D^2(z)$  is the average poverty gap for poverty line  $z$ . If, for all  $z \in [z_-, z_+]$ ,  $D_A^2(z) > D_B^2(z)$ , then it follows that the average poverty gap is greater in  $A$  than in  $B$  for all poverty lines in the interval  $[z_-, z_+]$ . But this condition is just restricted stochastic dominance of  $A$  by  $B$  over that interval.

First-order dominance implies second-order dominance, but second-order dominance does not necessarily imply first-order dominance. Thus second-order dominance, third-order dominance, fourth-order dominance, *etc.*, are progressively weaker conditions. We will now see that they imply unanimous rankings of two populations for progressively more and more restricted classes of poverty indices. Suppose that the function  $\pi$  is twice differentiable, with positive second derivative  $\pi''$ . This condition can perhaps be interpreted as increasing marginal (social) disutility of the poverty gap. Then integrating once more by parts gives

$$\Pi_A(z) - \Pi_B(z) = [\pi'(z - y)(D_A^2(y) - D_B^2(y))]_0^z + \int_0^z \pi''(z - y)(D_A^2(y) - D_B^2(y)) dy.$$

With  $\pi' > 0$  and  $\pi'' > 0$ , this expression is positive if the condition for second-order stochastic dominance is satisfied. Thus second-order dominance is sufficient (and necessary) for all indices with  $\pi' > 0$  and  $\pi'' > 0$  to be unanimous in their ranking.

We can go further, and we find that indices for which  $\pi^{(s)} \geq 0$ ,  $\pi^{(s-1)}(0) \geq 0$ , and  $\pi^{(i)}(0) = 0$  for  $i = 1, \dots, s - 2$  rank unanimously if and only if there is stochastic dominance at order  $s$ .

## Lorenz Dominance

The concept of **Lorenz dominance** is easy to define. A distribution  $B$  Lorenz dominates another distribution  $A$  if  $L_B(p) \geq L_A(p)$  for all  $p \in [0, 1]$ . The inequality goes in the other direction than that used in the definition of stochastic dominance, because the Lorenz curve for  $B$  is closer to the 45-degree line than that of  $A$ , which means that there is less inequality in  $B$ . As remarked earlier, Lorenz curves do not respect the Pareto principle, and so neither does Lorenz dominance.

This means that, although Lorenz dominance is fine for comparisons of inequality and only that, it is not suitable for welfare comparisons, and, in particular, for poverty comparisons. This defect is remedied by the concept of **generalised Lorenz dominance**, based on a generalised Lorenz curve. The generalisation simply removes the mean income  $\mu$  from the definition of the ordinate. We have

$$GL(p) = \int_0^{Q(p)} y \, dF(y),$$

and distribution  $B$  dominates  $A$  is  $GL_B(p) \geq GL_A(p)$  for all  $p \in [0, 1]$ .

It turns out that generalised Lorenz dominance is the same thing as second-order stochastic dominance. In order to see this, note that the condition for second-order dominance of  $A$  by  $B$ , which is  $D_A^2(y) \geq D_B^2(y)$ , can be written as

$$\int_0^y (F_A(x) - F_B(x)) dx \geq 0 \text{ for all } y > 0.$$

This follows from the definition of  $D^2$  as the integral of the CDF  $F$ . Generalised Lorenz dominance, on the other hand, can be rewritten by making a change of variable in the definition of  $GL(p)$ . If we put  $q = F(y)$ , then, when  $y = 0$ ,  $q = 0$ , and when  $y = Q(p)$ ,  $q = F(Q(p)) = p$ , since the CDF  $F$  and the quantile function  $Q$  are inverse functions. Thus

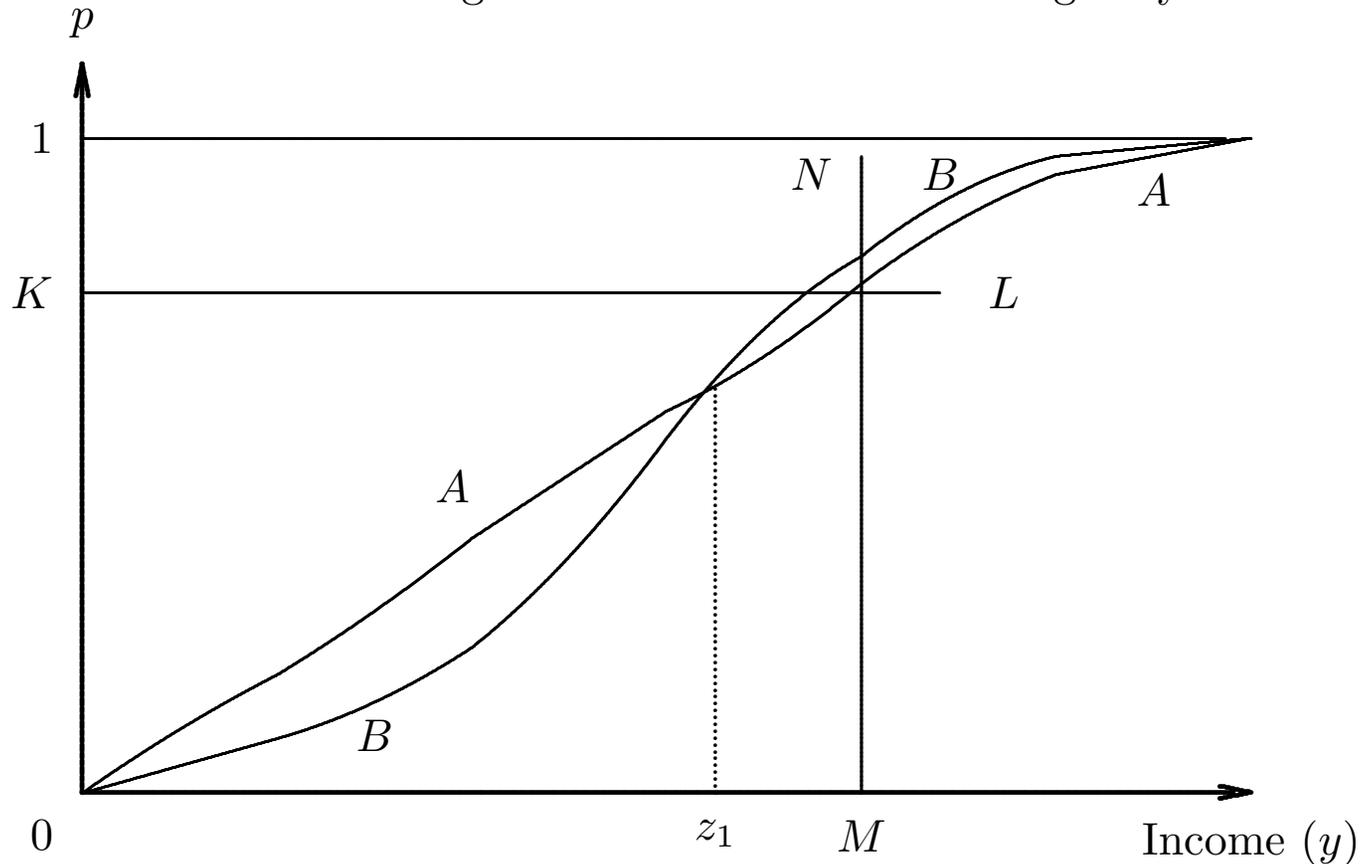
$$GL(p) = \int_0^p Q(q) dq,$$

and so generalised Lorenz dominance can be written as

$$\int_0^p (Q_B(q) - Q_A(q)) dq \geq 0 \text{ for all } p \in [0, 1].$$

We have to show, then, that the conditions are equivalent.

Consider the setup below. Here distribution  $B$  dominates  $A$  at second order because, although the CDFs cross, the areas between them are such that the condition for second-order dominance is always satisfied. Thus the vertical line  $MN$  marks off a large positive area between the graphs of the two CDFs up to the point at which they cross, and thereafter a small negative area bounded on the right by  $MN$ .



**Generalised Lorenz and Second Order Dominance**

Now, for generalised Lorenz dominance, the integral

$$\int_0^p (Q_B(q) - Q_A(q)) dq$$

is the area between the two curves, interpreted this time as the graphs of the quantile functions  $Q_A$  and  $Q_B$ , and thus bounded above by the horizontal line  $KL$ . Although it is tedious to demonstrate it algebraically, it is intuitively clear that if the areas bounded on the right by vertical lines like  $MN$  are always positive, then so are the areas bounded above by horizontal lines like  $KL$ , that is, integrals like

$$\int_0^y (F_A(x) - F_B(x)) dx.$$

This shows that the two kinds of dominance are indeed equivalent.

## Higher-Order Dominance

Another thing that emerges clearly from the Figure above is that the threshold income  $z_1$  up to which first-order stochastic dominance holds is always smaller than the threshold  $z_2$  up to which we have second-order dominance. In the Figure, of course, we have second-order dominance everywhere, and so we can set  $z_2$  equal to the highest income in either distribution. More generally, we can define a threshold  $z_s$  as the greatest income up to which we have dominance at order  $s$ :

$$z_s = \inf_y \{D_B^s(y) \geq D_A^s(y)\}.$$

Thus the  $z_s$  constitute an increasing sequence.

The following Lemma, taken from Davidson and Duclos (2000), shows that we can always find an order  $s$  such that there is dominance at order  $s$  all the way up to the highest income.

### Lemma 1:

If  $B$  dominates  $A$  for  $s = 1$  up to some  $w > 0$ , with strict dominance over at least part of that range, then for any finite threshold  $z$ ,  $B$  dominates  $A$  at order  $s$  up to  $z$  for  $s$  sufficiently large.

## Proof:

We have  $F_A(x) - F_B(x) \geq 0$  for  $0 \leq x \leq w$ , with strict inequality over some subinterval of  $[0, w]$ . Thus

$$\int_0^w (F_A(y) - F_B(y)) \, dy \equiv a > 0.$$

We wish to show that, for arbitrary finite  $z$ , we can find  $s$  sufficiently large that  $D_A^s(x) - D_B^s(x) > 0$  for  $x \leq z$ , that is,

$$\int_0^x \left(1 - \frac{y}{x}\right)^{s-1} (dF_A(y) - dF_B(y)) > 0$$

for  $x < z$ . For ease in the sequel, we have multiplied  $D^s(x)$  by  $(s-1)!/x^{s-1}$ , which does not affect the inequality we wish to demonstrate.

The left-hand side can be integrated by parts to yield

$$\frac{s-1}{x} \int_0^x (F_A(y) - F_B(y)) \left(1 - \frac{y}{x}\right)^{s-2} \, dy.$$

We split this integral in two parts: the integral from 0 to  $w$ , and then from  $w$  to  $x$ . We may bound the absolute value of the second part: Since  $|F_A(y) - F_B(y)| \leq 1$  for any  $y$  and  $1 - y/x \geq 0$  for all  $y \leq x$ , we have

$$\begin{aligned} & \left| \frac{s-1}{x} \int_w^x (F_A(y) - F_B(y)) \left(1 - \frac{y}{x}\right)^{s-2} dy \right| \\ & \leq \frac{s-1}{x} \int_w^x \left(1 - \frac{y}{x}\right)^{s-2} dy = \left(1 - \frac{w}{x}\right)^{s-1}. \end{aligned}$$

For the range from 0 up to  $w$ , we have, for  $s \geq 2$ ,

$$\begin{aligned} & \frac{s-1}{x} \int_0^w (F_A(y) - F_B(y)) \left(1 - \frac{y}{x}\right)^{s-2} dy \\ & \geq \frac{s-1}{x} \left(1 - \frac{w}{x}\right)^{s-2} \int_0^w (F_A(y) - F_B(y)) dy \\ & = \frac{a(s-1)}{x} \left(1 - \frac{w}{x}\right)^{s-2} \end{aligned}$$

Putting everything together, we find that, for  $x \geq w$ ,

$$\begin{aligned}
 & \int_0^x \left(1 - \frac{y}{x}\right)^{s-1} (\mathrm{d}F_A(y) - \mathrm{d}F_B(y)) \\
 & \geq \frac{a(s-1)}{x} \left(1 - \frac{w}{x}\right)^{s-2} - \left(1 - \frac{w}{x}\right)^{s-1} \\
 & = \left(1 - \frac{w}{x}\right)^{s-2} \left(\frac{a(s-1)}{x} - 1 + \frac{w}{x}\right).
 \end{aligned}$$

If we choose  $s$  to be greater than  $1 + (z - w)/a$ , then, for all  $w \leq x \leq z$ ,  $(a(s - 1) + w)/x - 1 > 0$ . Thus for such  $s$ , the last expression in the display above is positive for all  $w \leq x \leq z$ . For  $x < w$ , the dominance at first order up to  $w$  implies dominance at any order  $s > 1$  up to  $w$ . The result is therefore proved. ■

## Better Poverty Indices

Although we found that, in order to satisfy some of the axioms we wish to maintain, we had to restrict our SWFs to anonymous and additively separable ones, that is not enough to satisfy all of our desiderata. Foster, Greer, and Thorbecke (1984) – FGT henceforth – made considerable progress in defining a class of indices with desirable properties.

They begin by proposing the following index:

$$P(z, \mathbf{y}) = \frac{1}{nz^2} \sum_{i=1}^q g_i^2.$$

Here the poverty gap  $g_i$  is weighted by itself, and this already goes a long way towards satisfying not only Sen's axioms but also others we have discussed. Like Sen's own proposal, which we no longer care for, it can be seen to depend on the headcount ratio  $H$ , the income-gap ratio  $I$ , and a measure of inequality, which turns out to be the squared coefficient of variation:

$$C_p^2 \equiv \frac{1}{q} \sum_{i=1}^q (y_i - \mu_p)^2 / \mu_p^2.$$

FGT show that

$$P(z, \mathbf{y}) = H(I^2 + (1 - I)^2 C_p^2).$$

This is not exactly what BD proposed, but it is very similar, and shares the good properties that BD sought. To see why this result holds, recall that  $H = q/n$  and  $I = 1/(qz) \sum_{i=1}^q (z - y_i) = 1 - \mu_p/z$ . Therefore

$$\begin{aligned} H(I^2 + (1 - I)^2 C_p^2) &= \frac{q}{n} \left[ \frac{1}{q^2 z^2} \left( \sum_{i=1}^q (z - y_i) \right)^2 + \frac{\mu_p^2}{z^2 q} \sum_{i=1}^q \frac{(y_i - \mu_p)^2}{\mu_p^2} \right] \\ &= \frac{1}{nz^2} \left[ q(z - \mu_p)^2 + \sum_{i=1}^q y_i^2 - 2\mu_p^2 q + q\mu_p^2 \right] \\ &= \frac{1}{nz^2} \left[ \sum_{i=1}^q y_i^2 + qz^2 - 2qz\mu_p \right] = \frac{1}{nz^2} \sum_{i=1}^q (z - y_i)^2. \end{aligned}$$

- *Transfer Sensitivity*

A positive transfer from a poor individual with income  $y$  to a slightly richer poor individual increases poverty more the smaller is  $y$ .

The index we have just looked at does not satisfy this axiom, or at least does so only weakly. However, there is an obvious generalisation which leads to a whole class of indices, called the FGT indices. They are defined as follows:

$$P_{\alpha}(z, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^q \left( \frac{g_i}{z} \right)^{\alpha}, \quad \alpha \geq 0.$$

FGT prove the following theorem.

### **Theorem**

The index  $P_{\alpha}$  satisfies the Monotonicity axiom for  $\alpha > 0$ , the Transfer axiom for  $\alpha > 1$ , and the Transfer Sensitivity axiom for  $\alpha > 2$ .

## Connection with Stochastic Dominance

We can slightly modify the definition of the index  $P_\alpha$  to make it depend on a CDF  $F$  rather than a realised sample  $\mathbf{y}$ . We have

$$P_\alpha(z, F) = \int_0^z \left( \frac{z - y}{z} \right)^\alpha dF(y).$$

Then the indices defined as above for a sample could as well be defined as functions of the EDFs of the samples. This definition makes it clear that  $\alpha = 0$  gives the headcount ratio  $H$ , while  $\alpha = 1$  gives the income-gap ratio  $I$ .

The FGT indices define a partial ordering of income distributions. Then distribution  $F$  has less poverty than distribution  $G$  if, for all  $z$  in some specified set  $Z$  of conceivable poverty lines,  $P_\alpha(z, F) < P_\alpha(z, G)$ .

### Theorem

$P_\alpha(z, F) < P_\alpha(z, G)$  for all  $z \in Z$  if and only if there is restricted stochastic dominance of  $G$  by  $F$  over  $Z$  at order  $\alpha + 1$ .

### Proof:

Recall that

$$D_F^{\alpha+1}(z) = \frac{1}{\alpha!} \int_0^z (z - y)^\alpha dF(y).$$

The result follows on noting that  $P_\alpha(z, F) < P_\alpha(z, G)$  if and only if  $D_F^{\alpha+1}(z) < D_G^{\alpha+1}(z)$ .

## Inference for Stochastic Dominance

From the definition of the **dominance functions** for a CDF  $F$ :

$$D^s(z) = \frac{1}{(s-1)!} \int_0^z (z-y)^{s-1} dF(y)$$

we may form the plug-in estimator for a sample  $\{y_i\}$ ,  $i = 1, \dots, n$  by replacing  $F$  by the EDF  $\hat{F}$ , to obtain:

$$\hat{D}^s(z) = \frac{1}{(s-1)!} n^{-1} \sum_{i=1}^n (z - y_i)_+^{s-1}$$

where by  $x_+$  we mean  $\max(0, x)$ . It can be seen that  $\hat{D}^s(z)$  is an average of IID realisations. This allows us to use the LLN and the CLT to study the asymptotic properties of  $\hat{D}^s(z)$  in the same way as we did earlier for the EDF itself.

### Theorem 1 (DD 2000)

Let the joint population moments of order  $2s-2$  of  $Y^A$  and  $Y^B$  be finite. Then  $n^{1/2}(\hat{D}_K^s(x) - D_K^s(x))$  is asymptotically normal with expectation zero, for  $K = A, B$ , and with asymptotic covariance structure given by  $(K, L = A, B)$

$$\begin{aligned} & \lim_{n \rightarrow \infty} n \operatorname{cov}(\hat{D}_K^s(x), \hat{D}_L^s(z)) \\ &= \frac{1}{((s-1)!)^2} \mathbf{E} \left( (x - Y^K)_+^{s-1} (z - Y^L)_+^{s-1} \right) - D_K^s(x) D_L^s(z). \end{aligned}$$

### Proof:

For each distribution, the existence of the population moment of order  $s-1$  lets us apply the law of large numbers, thus showing that  $\hat{D}^s(x)$  is a consistent estimator of  $D^s(x)$ . Given also the existence of the population moment of order  $2s-2$ , the central limit theorem shows that the estimator is root- $n$  consistent and asymptotically normal with the asymptotic covariance matrix given above. This formula clearly applies not only for  $Y^A$  and  $Y^B$  separately, but also for the covariance of  $\hat{D}_A^s$  and  $\hat{D}_B^s$ .

If  $A$  and  $B$  are independent populations, the sample sizes  $n_A$  and  $n_B$  may be different. Then the given formula applies to each with  $n$  replaced by the appropriate sample size. The covariance across the two populations is of course zero. ■

The asymptotic covariance matrix given in the Theorem can readily be consistently estimated in a distribution-free manner by using sample equivalents. Thus  $D^s(x)$  is estimated by  $\hat{D}^s(x)$ , and the expectation in the formula by

$$\frac{1}{n} \sum_{i=1}^n (x - y_i^K)_+^{s-1} (z - y_i^L)_+^{s-1}. \quad (1)$$

If  $B$  does dominate  $A$  weakly at order  $s$  up to some possibly infinite threshold  $z$ , then, for all  $x \leq z$ ,  $D_A^s(x) - D_B^s(x) \geq 0$ . There are various hypotheses that could serve either as the null or the alternative in a testing procedure. The most restrictive of these, which we denote  $H_0$ , is that  $D_A^s(x) - D_B^s(x) = 0$  for all  $x \leq z$ . Next comes  $H_1$ , according to which  $D_A^s(x) - D_B^s(x) \geq 0$  for  $x \leq z$ , and, finally,  $H_2$ , which imposes no restrictions at all on  $D_A^s(x) - D_B^s(x)$ . We observe that these hypotheses are nested:  $H_0 \subset H_1 \subset H_2$ .

In Theorem 1, it was assumed that the arguments  $x$  and  $z$  of the functions  $D^s$  were non-stochastic. In applications, one often wishes to deal with  $D^s(z - x)$ , where  $z$  is the poverty line. In the next Theorem, we deal with the case in which  $z$  is estimated on the basis of sample information.

## Theorem 2 (DD 2000)

Let the joint population moments of order  $2s - 2$  of  $y^A$  and  $y^B$  be finite. If  $s = 1$ , suppose in addition that  $F_A$  and  $F_B$  are differentiable and let  $D^0(x) = F'(x)$ . Assume first that  $n$  independent drawings of pairs  $(y^A, y^B)$  have been made from the joint distribution of  $A$  and  $B$ . Also, let the poverty lines  $z_A$  and  $z_B$  be estimated by  $\hat{z}_A$  and  $\hat{z}_B$  respectively, where these estimates are expressible asymptotically as sums of IID variables drawn from the same sample, so that, for some function  $\xi_A(\cdot)$ ,

$$\hat{z}_A = n^{-1} \sum_{i=1}^n \xi_A(y_i^A) + o(1) \quad \text{as } n \rightarrow \infty,$$

and similarly for  $B$ . Then  $n^{1/2}(\hat{D}_K^s(\hat{z}_K - x) - D_K^s(z_K - x))$ ,  $K = A, B$ , is asymptotically normal with mean zero, and with covariance structure given by  $(K, L = A, B)$

$$\begin{aligned} \lim_{n \rightarrow \infty} n \operatorname{cov}(\hat{D}_K^s(\hat{z}_K - x), \hat{D}_L^s(\hat{z}_L - x')) = \\ \operatorname{cov}\left(D_K^{s-1}(z_K - x)\xi_K(Y^K) + ((s-1)!)^{-1}(z_K - x - Y^K)_+^{s-1}, \right. \\ \left. D_L^{s-1}(z_L - x')\xi_L(Y^L) + ((s-1)!)^{-1}(z_L - x' - Y^L)_+^{s-1}\right). \end{aligned}$$

If  $Y^A$  and  $Y^B$  are independently distributed, and if  $n_A$  and  $n_B$  IID drawings are respectively made of these variables, then, for  $K = L$ ,  $n_K$  replaces  $n$ , while for  $K \neq L$ , the covariance is zero.

The most popular choices of population dependent poverty lines are fractions of the population mean or median, or quantiles of the population distribution. Clearly any function of a sample moment can be expressed asymptotically as an average of IID variables, and the same is true of functions of quantiles, at least for distributions for which the density exists, according to the Bahadur representation of quantiles.

### **Lemma (Bahadur (1966))**

Let the CDF  $F$  be twice differentiable. Then, for the population quantile  $Q(p)$  and the sample quantile  $\hat{Q}(p)$  we have

$$n^{1/2}(\hat{Q}(p) - Q(p)) = -\frac{n^{-1/2}}{f(Q(p))} \sum_{i=1}^n (\mathbb{I}(y_i < Q(p)) - p) + O(n^{-3/4}(\log n)^{3/4}),$$

where  $f = F'$  is the density.

The result implies that  $\hat{Q}(p)$  is root- $n$  consistent, and that it can be expressed asymptotically as an average of IID variables. When the poverty line is a proportion  $k$  of the median, for instance, we have that:

$$\xi(y_i) = -k \left( \frac{\mathbf{I}(y_i < Q(0.5)) - 0.5}{F'(Q(0.5))} \right),$$

where  $Q(0.5)$  denotes the median. When  $z$  is  $k$  times average income, we have

$$\xi(y_i) = ky_i.$$

This IID structure makes it easy to compute asymptotic covariance structures for sets of quantiles of jointly distributed variables.

Assume that  $\hat{F}_A(x) > \hat{F}_B(x)$  for some bottom range of  $x$ . If  $\hat{F}_A(x) < \hat{F}_B(x)$  for larger values of  $x$ , define  $z_1$  as the smallest income for which  $\hat{F}_A(z_1) \leq \hat{F}_B(z_1)$ . A natural estimator of  $z_1$  is  $\hat{z}_1$  defined implicitly by

$$\hat{F}_A(\hat{z}_1) = \hat{F}_B(\hat{z}_1).$$

If  $\hat{F}_A(x) > \hat{F}_B(x)$  for all  $x \leq z$ , for some prespecified poverty line  $z$ , then we arbitrarily set  $\hat{z}_1 = z$ . If  $\hat{z}_1$  is less than the poverty line  $z$ , we may define  $\hat{z}_2$  by

$$\hat{D}_A^2(\hat{z}_2) = \hat{D}_B^2(\hat{z}_2)$$

if this equation has a solution less than  $z$ , and by  $z$  otherwise. And so on for  $\hat{z}_s$  for  $s > 2$ : either we can solve the equation

$$\hat{D}_A^s(\hat{z}_s) = \hat{D}_B^s(\hat{z}_s),$$

or else we set  $\hat{z}_s = z$ . Note that the second possibility is a mere mathematical convenience used so that  $\hat{z}_s$  is always well defined – we may set  $z$  as large as we wish. The following theorem gives the asymptotic distribution of  $\hat{z}_s$  under the assumption that  $z_s < z$  exists in the population.

### Theorem 3 ((DD 2000))

Let the joint population moments of order  $2s - 2$  of  $Y^A$  and  $Y^B$  be finite. If  $s = 1$ , suppose further that  $F_A$  and  $F_B$  are differentiable, and let  $D^0(x) = F'(x)$ . Suppose that there exists  $z_s < z$  such that  $D_A^s(z_s) = D_B^s(z_s)$ , and that  $D_A^s(x) > D_B^s(x)$  for all  $x < z_s$ . Assume that  $z_s$  is a simple zero, so that the derivative  $D_A^{s-1}(z_s) - D_B^{s-1}(z_s)$  is nonzero. If we have a sample of pairs from the joint distribution of  $(Y^A, Y^B)$ , then  $n^{1/2}(\hat{z}_s - z_s)$  is asymptotically normally distributed with expectation zero, and asymptotic variance given by:

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Var}(n^{1/2}(\hat{z}_s - z_s)) &= \left( (s-1)! (D_A^{s-1}(z_s) - D_B^{s-1}(z_s)) \right)^{-2} \times \\ &\left( \text{Var}((z_s - Y^A)_+^{s-1}) + \text{Var}((z_s - Y^B)_+^{s-1}) \right. \\ &\quad \left. - 2 \text{cov}((z_s - Y^A)_+^{s-1}, (z_s - Y^B)_+^{s-1}) \right). \end{aligned}$$

If  $Y^A$  and  $Y^B$  are independently distributed, and if the ratio  $r \equiv n_A/n_B$  remains constant as  $n_A$  and  $n_B$  tend to infinity, then  $n_A^{1/2}(\hat{z}_s - z_s)$  is asymptotically normal with mean zero, and asymptotic variance given by

$$\lim_{n_A \rightarrow \infty} \text{Var}(n_A^{1/2}(\hat{z} - z)) = \frac{\text{Var}((z - Y^A)_+^{s-1}) + r \text{Var}((z - Y^B)_+^{s-1})}{\left( (s-1)! (D_A^{s-1}(z) - D_B^{s-1}(z)) \right)^2}.$$

## Restricted Dominance and Non-Dominance

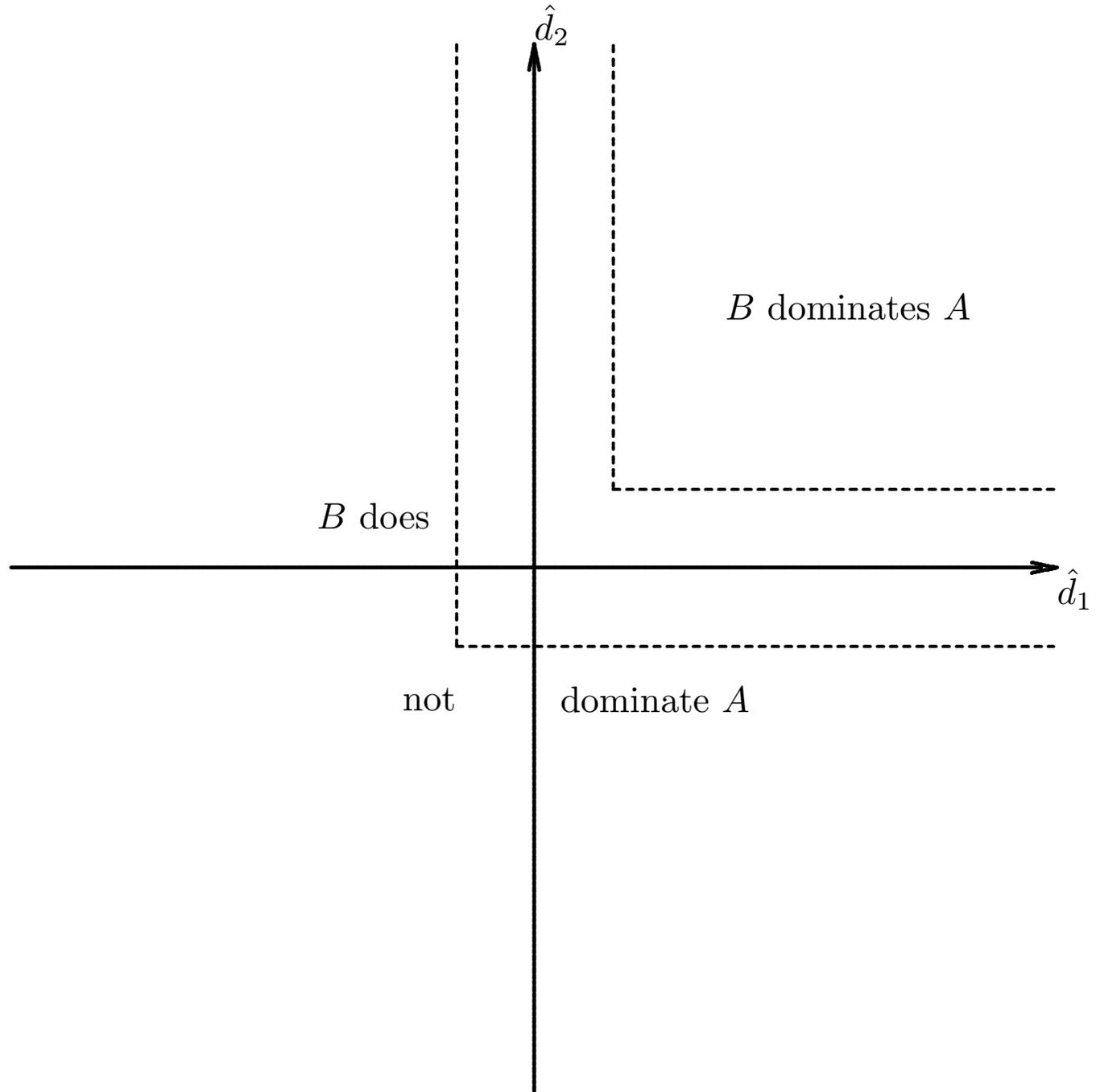
The most common approach to test whether there is stochastic dominance, on the basis of samples drawn from the two populations  $A$  and  $B$ , is to posit a null hypothesis of dominance, and then to study test statistics that may or may not lead to rejection of this hypothesis. This is arguably a matter of convention and convenience: convention in the sense that it follows the usual practice of making the theory of interest the null and seeking evidence contrary to it, and convenience in that the null is then relatively easy to formulate.

Rejection of a null of dominance can, however, sometimes be viewed as an inconclusive outcome since it fails to rank the two populations. Further, in the absence of information on the power of the tests, non-rejection of dominance does not enable one to accept dominance, which is nevertheless often the outcome of interest. Hence, under this first approach, stochastic dominance merely remains either contradicted or uncontradicted, but cannot be established.

From a logical point of view, it may thus seem desirable in some settings to posit instead a null of *non*-dominance. If we succeed in rejecting this null, we may indeed then legitimately infer the only other possibility, namely dominance.

In order to clarify the above, it may be useful to consider a very simple case with two distributions  $A$  and  $B$  with the same support, consisting of three points,  $y_1 < y_2 < y_3$ . Since  $F_A(y_3) = F_B(y_3) = 1$ , inference on stochastic dominance can be based on just two quantities,  $\hat{d}_i \equiv \hat{F}_A(y_i) - \hat{F}_B(y_i)$ , for  $i = 1, 2$ . The hats indicate estimates of the CDFs at the two points. Distribution  $B$  dominates distribution  $A$  if, in the population,  $d_i \geq 0$ .

The following Figure shows a two-dimensional plot of  $\hat{d}_1$  and  $\hat{d}_2$ . The first quadrant corresponds to dominance of  $A$  by  $B$  in the sample. In order to reject a hypothesis of dominance, therefore, the observed  $\hat{d}_1$  and  $\hat{d}_2$  must lie significantly far away from the first quadrant, for example, in the area marked as “ $B$  does not dominate  $A$ ” separated from the first quadrant by an L-shaped band. This is essentially the procedure followed by the first approach described above, which is based on testing a null of dominance.



For a rejection of non-dominance, on the other hand, the observed sample point must lie “far enough” inside the first quadrant that it is significantly removed from the area of nondominance, as in the area marked “ $B$  dominates  $A$ ”. The challenge is to assess what is “far enough”. The zone between the rejection regions for the two possible null hypotheses of dominance and non-dominance corresponds to situations in which neither hypothesis can be rejected. We see that this happens when one of the  $\hat{d}_i$  is close to zero and the other is positive. Note also from the Figure that inferring dominance by rejecting the hypothesis of non-dominance is more demanding than failing to reject the hypothesis of dominance, since, for dominance, both statistics must have the same sign and be statistically significant.

The two approaches described above can thus be seen as complementary. Positing a null of dominance cannot be used to infer dominance; it can however serve to infer non-dominance. Positing a null of non-dominance cannot serve to infer non-dominance; it can however lead to inferring dominance.

## The minimum $t$ statistic

In Kaur, Prakasa Rao, and Singh (1994) (KPS), a test is proposed based on the minimum of the  $t$  statistic for the hypothesis that  $F_A(z) - F_B(z) = 0$ , computed for each value of  $z$  in some closed interval contained in the interior of  $U$ , the union of the supports of  $F_A$  and  $F_B$ . The minimum value is used as the test statistic for the null of non-dominance against the alternative of dominance. The test can be interpreted as an intersection-union test. It is shown that the probability of rejection of the null when it is true is asymptotically bounded by the nominal level of a test based on the standard normal distribution. Howes (1993) proposed a very similar intersection-union test, except that the  $t$  statistics are calculated only for the predetermined grid of points.

## Empirical Likelihood

For a given sample, the “parameters” of the empirical likelihood are the probabilities associated with each point in the sample. The empirical loglikelihood function (ELF) is then the sum of the logarithms of these probabilities. Let  $Y$  be the set of distinct observations  $y_i$  in the sample, and let  $n_i$  denote the number of observations equal to  $y_i$ . If there are no constraints, the ELF is maximised by solving the problem

$$\max_{p_i} \sum_{y_i \in Y} n_i \log p_i \quad \text{subject to} \quad \sum_{y_i \in Y} p_i = 1.$$

It is easy to see that the solution to this problem is  $p_i = n_i/N$  for all  $i$ . The maximised ELF is  $-n \log n + \sum_i n_i \log n_i$ , an expression which has a well-known entropy interpretation.

With two samples,  $A$  and  $B$ , we see that the probabilities that solve the problem of the unconstrained maximisation of the total ELF are  $p_i^K = n_i^K / n_K$  for  $K = A, B$ , and that the maximised ELF is

$$-n_A \log n_A - n_B \log n_B + \sum_{y_i^A \in Y^A} n_i^A \log n_i^A + \sum_{y_i^B \in Y^B} n_i^B \log n_i^B.$$

Notice that, in the continuous case, and in general whenever  $n_i^K = 1$ , the term  $n_i^K \log n_i^K$  vanishes.

For an empirical likelihood-ratio test, the statistic, just as with ordinary likelihood, is twice the difference between the ELF maximised without constraint and the ELF maximised subject to the constraints of the null hypothesis. For given  $z$  we may test the same hypothesis as we could test with an asymptotic  $t$  statistic. Let  $n_A(z)$  be the number of observations in sample  $A$  less than or equal to  $z$  and let  $m_A(z) = n_A - n_A(z)$ , and similarly for  $n_B(z)$  and  $m_B(z)$ . It can be seen that the EL ratio statistic is twice

$$\begin{aligned} & n \log n - n_A \log n_A - n_B \log n_B + n_A(z) \log n_A(z) + n_B(z) \log n_B(z) \\ & + m_A(z) \log m_A(z) + m_B(z) \log m_B(z) - (n_A(z) + n_B(z)) \log(n_A(z) + n_B(z)) \\ & - (m_A(z) + m_B(z)) \log(m_A(z) + m_B(z)). \end{aligned}$$

Fortunately, the rather complicated ELR statistic is asymptotically equivalent to the much simpler  $t$  statistic.

### Theorem 1 (DD (2013))

As the size  $n$  of the combined sample tends to infinity in such a way that  $n_A/n \rightarrow r$ ,  $0 < r < 1$ , the difference between the statistic  $\text{LR}(z)$  and the squared  $t$  statistic  $t^2(z)$  is of order  $n^{-1/2}$  for any point  $z$  in the interior of  $U$ , the union of the supports of populations  $A$  and  $B$ , such that  $F_A(z) = F_B(z)$ .

### Corollary

Under local alternatives to the null hypothesis that  $F_A(z) = F_B(z)$ , where  $F_A(z) - F_B(z)$  is of order  $n^{-1/2}$  as  $n \rightarrow \infty$ , the asymptotic equivalence of  $t^2(z)$  and  $\text{LR}(z)$  continues to hold.

## The tails of the distributions

Although the null of nondominance has the attractive property that, if it is rejected, all that is left is dominance, this property comes at a cost, which is that it is impossible to infer dominance over the full support of the distributions if these distributions are continuous in the tails.

The non-dominance of distribution  $A$  by  $B$  implies that  $\max_{z \in U} (F_B(z) - F_A(z)) \geq 0$ . But if  $z^-$  denotes the lower limit of  $U$ , we must have  $F_B(z^-) - F_A(z^-) = 0$ , whether or not the null is true. Thus the maximum over the whole of  $U$  is never less than 0. Rejecting the null by a statistical test is therefore impossible.

Of course, an actual test is carried out, not over all of  $U$ , but only at the elements of the set  $Y$  of points observed in one or other sample. Suppose that  $A$  is dominated by  $B$  in the sample. Then the smallest element of  $Y$  is the smallest observation,  $y_1^A$ , in the sample drawn from  $A$ . The squared  $t$  statistic for the hypothesis that  $F_A(y_1^A) - F_B(y_1^A) = 0$  is then

$$t_1^2 \equiv \frac{n_A n_B (\hat{F}_A^1 - \hat{F}_B^1)^2}{n_B \hat{F}_A^1 (1 - \hat{F}_A^1) + n_A \hat{F}_B^1 (1 - \hat{F}_B^1)},$$

where  $\hat{F}_K^1 = \hat{F}_K(y_1^A)$ ,  $K = A, B$ . Now  $\hat{F}_B^1 = 0$  and  $\hat{F}_A^1 = 1/n_A$ , so that

$$t_1^2 = \frac{n_A n_B / n_A^2}{(n_B / n_A)(1 - 1/n_A)} = \frac{n_A}{n_A - 1}.$$

The  $t$  statistic itself is thus approximately equal to  $1 + 1/(2n_A)$ . Since the minimum over  $Y$  of the  $t$  statistics is no greater than  $t_1$ , and since  $1 + 1/(2n_A)$  is nowhere near the critical value of the standard normal distribution for any conventional significance level, it follows that rejection of the null of non-dominance is impossible.

If the data are discrete or censored in the tails, it is no longer impossible to reject the null if there is enough probability mass in the atoms at either end or over the censored areas of the distribution. If the distributions are continuous but are discretised or censored, then it becomes steadily more difficult to reject the null as the censoring becomes less severe, and in the limit once more impossible. The difficulty is clearly that, in the tails of continuous distributions, the amount of information conveyed by the sample tends to zero, and so it becomes impossible to discriminate among different hypotheses about what is going on there. Focussing on *restricted* stochastic dominance is then the only empirically sensible course to follow.

## Testing

An important preliminary remark is that, if there is non-dominance in the sample, then there is no need for a formal test, since there is no question of rejecting the null of non-dominance.

For the remainder of our discussion, therefore, we restrict the null hypothesis to the **frontier** of non-dominance, that is, to distributions such that  $F_A(z_0) = F_B(z_0)$  for exactly one point  $z_0$  in  $[z^-, z^+]$ , and  $F_A(z) > F_B(z)$  with strict inequality for all  $z \neq z_0$  in that interval. These distributions constitute the **least favourable case** of the hypothesis of non-dominance in the sense that, with either the minimum  $t$  statistic or the minimum EL statistic, the probability of rejection of the null is no smaller on the frontier than with any other configuration of nondominance.

### Theorem 3 (DD 2013)

The minima over  $z$  of both the signed asymptotic  $t$  statistic  $t(z)$  and the signed empirical likelihood ratio statistic  $\text{LR}^{1/2}(z)$  are asymptotically pivotal for the null hypothesis that the distributions  $A$  and  $B$  lie on the frontier of restricted non-dominance of  $A$  by  $B$ .

For configurations that lie on the frontier, the asymptotic distribution of both statistics is  $N(0,1)$ . Use of the quantiles of this distribution as critical values for the test leads to an asymptotically conservative test when there is non-dominance inside the frontier.

## The Bootstrap

The fact that the statistics are asymptotically pivotal means that we can use the **bootstrap** to perform tests that should benefit from asymptotic refinements in finite samples; see Beran (1988). Specifically, the difference between the true rejection probability under the null hypothesis and the nominal level of the tests should converge to zero faster when the bootstrapped statistic is asymptotically pivotal than otherwise.

The reason for spending time on an ELR test that is asymptotically equivalent to a  $t$  test is that it provides empirical distributions that satisfy the constraints of the null hypothesis. The constrained empirical-likelihood estimates of the CDFs of the two distributions  $K = A, B$  can be written as

$$\tilde{F}_K(z) = \sum_{y_i^K \leq z} p_i^K n_i^K k,$$

In any implementation of a bootstrap test, the **bootstrap DGP** used to generate bootstrap samples must satisfy the null hypothesis under test. We may use the constrained EL estimates above in a procedure of **weighted resampling** in order to draw bootstrap samples. The **bootstrap  $P$  value** is then the proportion of the bootstrap  $t$  statistics that are farther out in the tail of the null distribution, as estimated by the bootstrap, than the  $t$  statistic computed from the original data.

## The Gini Index Revisited

Our previous discussion of the Gini index made no mention of statistical inference. It is only recently that reliable methods of inference for Gini indices were developed. But the principles that underlie these methods have all been studied already in this course. The idea is as usual to express the Gini as asymptotically equivalent to an average of IID realisations, and then to apply the LLN and the CLT to demonstrate asymptotic normality and derive an expression for the asymptotic covariance, which can then be estimated in a distribution-free manner.

The natural plug-in estimator,  $\hat{G}$ , is

$$\hat{G} = \frac{2}{\hat{\mu}} \int_0^{\infty} y \hat{F}(y) d\hat{F}(y) - 1.$$

We discussed the ambiguity that arises from the cadlag definition of  $\hat{F}$ , and resolved it by splitting the difference, to get

$$\hat{G} = \frac{1}{\hat{\mu}n^2} \sum_{i=1}^n y_{(i)}(2i - 1) - 1.$$

## An asymptotic expression

Let

$$I \equiv \int_0^{\infty} yF(y) dF(y) \quad \text{and} \quad \hat{I} \equiv \int_0^{\infty} y\hat{F}(y) d\hat{F}(y).$$

Then we have

$$\begin{aligned} n^{1/2}(\hat{G} - G) &= n^{1/2} \left( \frac{2\hat{I}}{\hat{\mu}} - \frac{2I}{\mu} \right) = n^{1/2} \frac{2}{\mu\hat{\mu}} (\mu\hat{I} - \hat{\mu}I) \\ &= \frac{2}{\mu\hat{\mu}} (\mu n^{1/2}(\hat{I} - I) - I n^{1/2}(\hat{\mu} - \mu)). \end{aligned}$$

Our assumed regularity ensures that both  $n^{1/2}(\hat{\mu} - \mu)$  and  $n^{1/2}(\hat{I} - I)$  are of order 1 in probability. To leading order, then, we may approximate by replacing  $\mu\hat{\mu}$  in the denominator by  $\mu^2$ .

Next, we note that

$$n^{1/2}(\hat{\mu} - \mu) = n^{-1/2} \sum_{j=1}^n (y_j - \mu).$$

Clearly this is an asymptotically normal random variable. Next we wish to show that  $n^{1/2}(\hat{I} - I)$  is so as well.

We calculate as follows.

$$\begin{aligned}
n^{1/2}(\hat{I} - I) &= n^{1/2} \left( \int_0^\infty y \hat{F}(y) d\hat{F}(y) - \int_0^\infty y F(y) dF(y) \right) \\
&= n^{1/2} \left( \int_0^\infty y F(y) d(\hat{F} - F)(y) + \int_0^\infty y(\hat{F}(y) - F(y)) dF(y) \right. \\
&\quad \left. + \int_0^\infty y(\hat{F}(y) - F(y)) d(\hat{F} - F)(y) \right).
\end{aligned}$$

The last term above is of order  $n^{-1/2}$  as  $n \rightarrow \infty$ , and so will be ignored for the purposes of our asymptotic approximation.

Note that  $I = E(YF(Y))$ . The first term in the rightmost member above is

$$n^{1/2} \int_0^\infty y F(y) d(\hat{F} - F)(y) = n^{-1/2} \sum_{j=1}^n (y_j F(y_j) - I);$$

Evidently, this is asymptotically normal, since the terms are IID, the expectation of each term in the sum is 0, and the variance exists. The second term is

$$n^{-1/2} \sum_{j=1}^n \left( \int_0^\infty y \mathbf{I}(y_j \leq y) dF(y) - I \right),$$

Define the deterministic function  $m(y) \equiv \int_0^y x \, dF(x)$ . We see that

$$\begin{aligned} \mathbf{E}(m(Y)) &= \int_0^\infty m(y) \, dF(y) = \int_0^\infty \int_0^y x \, dF(x) \, dF(y) \\ &= \int_0^\infty x \int_x^\infty dF(y) \, dF(x) = \int_0^\infty x(1 - F(x)) \, dF(x) \\ &= \mathbf{E}(Y(1 - F(Y))) = \mu - I. \end{aligned}$$

Consequently,

$$\begin{aligned} \int_0^\infty y \mathbf{I}(y_j \leq y) \, dF(y) - I &= \int_{y_j}^\infty y \, dF(y) - I \\ &= \mu - m(y_j) - I = -\left(m(y_j) - \mathbf{E}(m(Y))\right). \end{aligned}$$

Thus the second term becomes

$$-n^{-1/2} \sum_{j=1}^n \left(m(y_j) - \mathbf{E}(m(Y))\right),$$

which is again asymptotically normal.

It follows that  $n^{1/2}(\hat{I} - I)$  is also asymptotically normal, and so

$$\begin{aligned} n^{1/2}(\hat{I} - I) &= n^{-1/2} \sum_{j=1}^n \left( y_j F(y_j) - m(y_j) - \mathbb{E}(Y F(Y) - m(Y)) \right) \\ &= n^{-1/2} \sum_{j=1}^n \left( y_j F(y_j) - m(y_j) - (2I - \mu) \right). \end{aligned}$$

Finally, we obtain an approximate expression for  $n^{1/2}(\hat{G} - G)$ :

$$n^{1/2}(\hat{G} - G) \approx -\frac{2}{\mu^2} I n^{1/2}(\hat{\mu} - \mu) + \frac{2}{\mu} n^{1/2}(\hat{I} - I)$$

This expression can be regarded as resulting from the application of the delta method. It is useful to express it as the sum of contributions from the individual observations, as follows:

$$n^{1/2}(\hat{G} - G) \approx n^{-1/2} \frac{2}{\mu} \sum_{j=1}^n \left( -\frac{I}{\mu} (y_j - \mu) + y_j F(y_j) - m(y_j) - (2I - \mu) \right)$$

In this way,  $n^{1/2}(\hat{G} - G)$  is expressed approximately as the normalised sum of a set of IID random variables of expectation zero, so that asymptotic normality is an immediate consequence. Since  $G = 2I/\mu - 1$ , the variance of the limiting distribution of  $n^{1/2}(\hat{G} - G)$  is

$$\frac{1}{\mu^2} \text{Var}\left(- (G + 1)Y + 2(YF(Y) - m(Y))\right).$$

### Estimating the variance

The next step is to see how to estimate the limiting variance in a distribution-free manner. First, we can estimate  $G$  by  $\hat{G}$  and  $\mu$  by  $\hat{\mu}$ . But the functions  $F$  and  $m$  are normally unknown, and so they, too, must be estimated. The value of  $F(y_{(i)})$  at the order statistic  $y_{(i)}$  is estimated by  $\hat{F}(y_{(i)}) = (2i - 1)/(2n)$ , where we continue to evaluate  $\hat{F}$  at its points of discontinuity by the average of the lower and upper limits. Since by definition  $m(y) = E(Y I(Y \leq y))$ , we can estimate  $m(y_j)$  by

$$\hat{m}(y_j) = \hat{E}(Y I(Y \leq y_j)) = \frac{1}{n} \sum_{i=1}^n y_i I(y_i \leq y_j).$$

If  $y_j = y_{(i)}$ , then we see that  $\hat{m}(y_{(i)}) = (1/n) \sum_{j=1}^i y_{(j)}$ .

Let  $Z$  be the random variable  $-(G + 1)Y + 2(YF(Y) - m(Y))$  of which the variance appears in the limiting variance of  $n^{1/2}(\hat{G} - G)$ . Let  $Z_i \equiv -(G + 1)y_{(i)} + 2(y_{(i)}F(y_{(i)}) - m(y_{(i)}))$ . Clearly, we can estimate  $Z_i$  by

$$\hat{Z}_i \equiv -(\hat{G} + 1)y_{(i)} + \frac{2i - 1}{n}y_{(i)} - \frac{2}{n} \sum_{j=1}^i y_{(j)}.$$

Then  $\bar{Z} \equiv n^{-1} \sum_{i=1}^n \hat{Z}_i$  is an estimate of  $E(Z)$ , and  $n^{-1} \sum_{i=1}^n (\hat{Z}_i - \bar{Z})^2$  is an estimate of  $\text{Var}(Z)$ . Since the  $Z_i$  are the IID realisations of  $Z$ , the variance of  $\hat{G}$  can be estimated by

$$\widehat{\text{Var}}(\hat{G}) = \frac{1}{(n\hat{\mu})^2} \sum_{i=1}^n (\hat{Z}_i - \bar{Z})^2.$$

## The bootstrap

Now that we have an estimate of the variance of  $\hat{G}$ , we have a standard error, and can form asymptotically pivotal statistics in order to test hypotheses about  $G$ . This means that we can get the asymptotic refinements of the bootstrap described in Beran (1988).

Specifically, in order to test the hypothesis that the population value of the Gini index is  $G_0$ , one first computes the statistic  $\tau \equiv (\hat{G} - G_0) / \hat{\sigma}_G$ , where the standard error  $\hat{\sigma}_G$  is the square root of the variance estimate. Then one generates  $B$  bootstrap samples of size  $n$  by resampling with replacement from the observed sample (assumed to be also of size  $n$ ). For bootstrap sample  $j$ , one computes a bootstrap statistic  $\tau_j^*$ , in exactly the same way as  $\tau$  was computed from the original data, but with  $G_0$  replaced by  $\hat{G}$ , in order that the hypothesis tested should be true of the bootstrap data-generating process. The bootstrap  $P$  value is then the proportion of the  $\tau_j^*$  that are more extreme than  $\tau$ . For a test at significance level  $\alpha$ , rejection occurs if the bootstrap  $P$  value is less than  $\alpha$ . For such a test, it is also desirable to choose  $B$  such that  $\alpha(B + 1)$  is an integer.

Bootstrap confidence intervals can also be based on the empirical distribution of the bootstrap statistics  $\tau_j^*$ . For an interval at nominal confidence level  $1 - \alpha$ , one estimates the  $\alpha/2$  and  $1 - \alpha/2$  quantiles of the empirical distribution, normally as the  $\lceil \alpha B/2 \rceil$  and  $\lceil (1 - \alpha/2)B \rceil$  order statistics of the  $\tau_j^*$ . Here  $\lceil \cdot \rceil$  denotes the ceiling function:  $\lceil x \rceil$  is the smallest integer not smaller than  $x$ . Let these estimated quantiles be denoted as  $q_{\alpha/2}$  and  $q_{1-\alpha/2}$  respectively. Then the bootstrap confidence interval is constructed as  $[\hat{G} - \hat{\sigma}_G q_{1-\alpha/2}, \hat{G} - \sigma_G q_{\alpha/2}]$ . It is of the sort referred to as a percentile- $t$ , or bootstrap- $t$ , confidence interval.

In order to test a hypothesis that the Gini indices are the same for two populations from which two independent samples have been observed, a suitable test statistic is  $(\hat{G}_1 - \hat{G}_2) / \sqrt{\hat{\sigma}_{G1}^2 + \hat{\sigma}_{G2}^2}$ . For each bootstrap repetition, a bootstrap sample is generated by resampling with replacement from each of the two samples, and then the bootstrap statistic is computed as  $(G_1^* - G_2^* - \hat{G}_1 + \hat{G}_2) / \sqrt{(\sigma_{G1}^*)^2 + (\sigma_{G2}^*)^2}$  in what should be obvious notation. If the samples are correlated, the denominator of the statistic should take account of the covariance, which can be estimated using the same formula as for the variance. Bootstrap samples are then generated by resampling pairs of observations.