Is the Gini index of inequality overly sensitive to changes in the middle of the income distribution?

By Joseph L. Gastwirth

Department of Statistics, George Washington University

Abstract: The Gini index is the most commonly used measure of income inequality. Like any single summary measure of a set of data it cannot capture all aspects that are of interest to researchers. One of its widely reported flaws is that it is supposed to be overly sensitive to changes in the middle of the distribution. This claim is examined by studying the effect of small transfers between households or an additional increment in income going to one member of the population on the value of the index. It turns out that the difference in the rank order of donor and recipient is usually the most important factor determining the change in the Gini index due to the transfer, which implies that transfers from an upper income household to a low income household receive more weight that transfers involving the middle. Transfers between two middle income households do affect a higher fraction of the population than other transfers but those transfers do not receive an excessive weight relative to other transfers because the difference in the ranks of donor and recipient is smaller than the corresponding difference in other transfers. Thus, transfers between two households in the middle of the distribution do not receive more weight than a transfer of the same amount from an upper income household to one in the lower part of the distribution. Similarly, the effect on the Gini index when a household in either tail of the distribution receives an additional increment is larger than when a middle income household receives it. Contrary to much of the literature, these results indicate that the Gini index is not overly sensitive to changes in the middle of the distribution. Indeed, it is more sensitive to changes in the lower and upper parts of the distribution than those in the middle.

1. Introduction

In his seminal article developing the relationship between measures of income inequality and an underlying social welfare function, Atkinson (1970, p. 255) noted that the Gini index was one of three measures that were sensitive to transfers at all income levels. After analysing the effect of an "infinitesimal" transfer of income from a household to one with lower income, he concludes (on p.256) that the Gini index gives more weight to transfers in the centre of the distribution than at the tails, i.e. the Gini index attaches more weight to transfers affecting the middle class. Allison (1978) provided a thorough review of the properties of several commonly used measures of inequality. After examining the effect of a transfer of an amount a from the household with the i^{th} largest income to the j^{th} largest, he concludes that for a typically shaped income distribution the Gini index tends to be most sensitive to transfers around the middle of the least sensitive to transfers among the very rich and very poor. Jasso (1979) pointed out that the formula for the effect of a transfer in Alison (1978) did not consider the possibility that a transfer would change the order, so the formula is only valid when the rank-order of the households is unaffected by the transfer (1979). Neither Jasso (1979) nor Alison (1979) comment on the effect of this on the conclusion that the Gini index is most sensitive to transfers in the middle of the distribution. Since the 1970's many authors (Ahn, 1997; Jones and Weinberg, 2000; Madden, 2000 p. 76; Boorghi, 2005; DeMaio, 2007; Cobham and Sumner, 2013; OECD, 2013; Pressman, 2013; Schmid and Stein, 2013; Chang, 2014; Bird and Zolt, 2015;Thewissen et al., 2015) have noted the sensitivity of the Gini index to changes in the middle of the distribution or to income differences around the mode (Jenkins, 2009). Others state that it is relatively insensitive to changes in the top and bottom part of the distribution and more sensitive to changes in the middle (Roberts and Willits, 2015). Referring to the Gini index, Green et al (1994, p. 59) write "An increase or decrease in the middle of the distribution will have a greater impact on the index than a similar change at either end, since there are more earners in the middle ranks." Some (Krozer, 2014; Cobham et al., 2015) and the Wikipedia entry on income inequality metrics describe it as being overly sensitive to changes in the middle of the distribution. The guide by the Australian Bureau of Statistics (2015) states "The Gini coefficient is sometimes criticised as being too sensitive to relative changes around the middle of the income distribution. This sensitivity arises because the derivation of the Gini coefficient reflects the ranking of the population, and ranking is most likely to change at the densest part of the income distribution, which is likely to be around the middle of the distribution". In their analysis of the growth in inequality in France, Fremeaux and Piketty (2014) observe that "the increase in income inequality during the 2000s is sharper for indices sensitive to the middle of the distribution like the Gini coefficient."

The purpose of this note is to re-examine this claim. For a typical income distribution with a density that first increases, reaches its mode and then decreases, the effect of a change in the distribution, due to a transfer or addition to the income of one household, will be seen to depend on whether it preserves or changes the order or ranks of the households, the difference between the donor and recipient of a transfer and who receives the additional income, which also increases the mean. In the case of a mean preserving transfer, one from the highest income recipient to the lowest has a *larger* impact on the value of the Gini index than a transfer of the same amount from any other household to one with less income. Transfers in the middle of the distribution, especially around the mode, however, will change the relative ranking of a *higher proportion* of the population but this does *not* necessarily imply that such a transfer has the largest impact on the numerical value of the Gini index because the change in its numerator also depends on the difference in the ranks of the donor and recipient. When *one* household receives an additional amount of income and the incomes of all others is unchanged, the number of households the recipient passes over to reach its higher rank does affect the Gini index but in a manner that gives slightly *less* weight to changes in the middle.

The transfers of primary interest in economics obey the Pigou-Dalton criteria, which state that transfers from a poorer to a richer household increase inequality, while transfers from a richer household to a poorer on that does not reverse their relative ranking decreases inequality. While the paper will emphasize these transfers, in a few illustrative examples this principle will not hold.

Several useful representations of the Gini index are reviewed in Section 2. One of them will be used in Section 3 to examine the effect on the Gini index of a small *order-preserving* transfer, e.g. an infinitesimal transfer from a high income household to one with less income, discussed by Atkinson (1970) or a small additional increment given to one household. In the first situation the effect of a transfer depends on the difference between the rank of the donor and the rank of the recipient; the largest decrease in the Gini index occurring when the highest ranked household transfers the small amount to the lowest ranked. The effect of one household receiving a small increment in income, which does not change the rank-order, depends on the rank of the recipient and again the Gini index decreases (increases) most when the poorest (richest) household receives the small increment. Section 4 examines the effect of a transfer of income that does *not preserve the order*, e.g. the recipient of the transfer now has more income than several households whose incomes previously were greater than the recipient and the rank of the donor might also decline. In this situation, the rank order of a larger fraction of the population around a middle income recipient or donor will change when that household is involved in the transfer; however, the change in the Gini index also depends on the difference in the ranks of the donor and recipient. The relative weight of these two components depends on the magnitude of this difference in ranks. When both the recipient and donor are in the middle of the distribution, the component of the change in the Gini index due to the number of households affected is relatively more important than the difference in the ranks, which should be small. When the donor is in the upper income region, the difference in after tax ranks will be more important. The trade-off in the relative weights of the two components of change in the value of the Gini index indicates that the Gini index reflects the effect of non-order preserving transfers in various parts of the distribution and is not overly sensitive to those involving the middle. Section 5 focuses on how the Gini index changes when there is a small increase in the total income, which is all given to *one* household. It turns out that the Gini index decreases (increases) the most when the increment is given to the lowest (highest) income recipient while the magnitude of the change is smaller when the recipient is in the middle of the distribution. The implications of the results are discussed in section 6.

2. Formulas for the Gini index

Although the Gini index of an income distribution is often defined in terms of the ratio of the area between the line of equality and the Lorenz curve, it will be more convenient to use the fact that it is the ratio of the mean difference (Δ) to twice the mean (μ).The estimate (md) of the mean difference is the average of the absolute value of the all pairs of incomes. For a sample $x_1,\ldots,x_n$ of observations from the population of interest, it is defined Kendall and Stuart, 1977, p.46) as:

$$md = \left(\frac{1}{n(n-1)}\right) \sum_{i,j=1}^{i,j=n} |x_i - x_j| \qquad (1)$$

Sometimes n(n-1) is replaced by n^2, but formula (1), which excludes comparisons of an observation with itself is commonly used to estimate Δ (Sudheesh and Dewan, 2013). The Gini index, G, of the distribution underlying the data is Δ/2μ. Denoting the sample mean by $\overline{x}$, the estimate of G is g = $md/(2\overline{x})$.

Thon (1982) reviews several expressions for G and g. To assess the impact of transfers and changes, a convenient form first orders the observations by their size, i.e. $x_1<x_{2<\ldots<}x_n$ so

$$g=[\sum_{i=1}^{n}(2i-n-1)x_i]/\ \bar{x}n(n-1). \qquad (2)$$

The numerator of (2) is one-half the md (David, 1968; David and Nagaraja, 2003) and gives weight (2i-n-1)/ $\bar{x}$n(n-1) to the i[th] observation. In the numerator, the largest observation, x_n, receives weight (n-1) and the smallest receives weight –(n-1)/ $\bar{x}$n(n-1). Indeed, starting from the first, the weight given to each successive order statistic increases by 2/$\bar{x}$n(n-1). If n=2m-1, so the median is the m[th] largest observation, it receives weight, *zero*. When n=2m, the median is the average of x_m and x_{m+1} and the numerator of (2) gives weight -1 to x_m and +1 to x_{m+1}. Thus, g gives *more* weight to both extremes and *less* weight to the observations in the middle observations. The mean, in the denominator weights each observation equally, so the relative weight given to each of the ordered observations *increases* with their distance from the median. Further properties and a wide variety of statistical applications of the mean difference and Gini index are discussed by Yitzhaki (2002) and Yitzhaki and Schectman (2013).

3. The effect of a small order preserving transfer or additional increment on the Gini index

Following Atkinson's (1970) examination of the effect of an "infinitesimal" transfer from a higher income recipient to a lower one, consider the effect of a small transfer of ε from x_j to x_i, where $x_i <x_j$, which is *not* large enough to change the order of the observations. The mean, $\bar{x}$ is unchanged and only the terms in the numerator of (2) that will change involve x_i and x_j. Thus, the change in the numerator is:

$$(2j-n-1)(x_j-\varepsilon) - (2i-n-1)(x_i+\varepsilon) = -2(j-i)\varepsilon, \qquad (3)$$

implying that g is changed by is -2(j-i)ε/$\bar{x}$n(n-1). In particular, the decrease in g due to a transfer of ε from any observation to the one immediately below, i.e. j-i=1 is -2ε/$\bar{x}$n(n-1); so the effect of such a transfer is the *same* throughout the distribution. From (3) it is clear that the largest decrease occurs when the highest income recipient transfers ε to the lowest and the magnitude of the change depends only on ε and twice the difference in the ranks of the donor and recipient. An order-preserving transfer from the highest income holder to the median receives one-half the weight as a transfer to the poorest. These considerations demonstrate that the Gini index is not overly sensitive to small order preserving transfers in the middle of the distribution as transfers to or from the middle do not have as large an effect on g as transfers from the upper end to the lower end of the distribution.

Next, consider the situation where a *order preserving* increase in the total income of size ε that it is given to one unit, e.g. the j[th] so x_j becomes x_j + ε and the mean becomes (n-1)(n$\bar{x}$ + ε), where $\bar{x}$ is the mean of the original observations. From formula (2) it follows that the Gini index (g_1) of the new data is:

$$g_1 = \frac{\sum_{i=1}^{n}(2i-n-1)x_i + \varepsilon(2j-n-1)}{(n-1)(n\bar{\boldsymbol{x}}+\boldsymbol{\epsilon})}. \tag{4}$$

The part of formula (4) that depends primarily on j is the second term in the numerator, which increases linearly in j. When j=1, i.e. the small additional income goes to the poorest household, the numerator *decreases* by (n-1)ε; when n=2j-1, so the j^{th} ranked member of the population is the median this term is zero and when j=n, i.e. the richest household receives the additional income, the numerator increases by (n-1)ε. Since the increase in the denominator of (4) is the same regardless of which household receives the increase, the *largest decline* in the Gini index occurs when j=1, i.e. the poorest household receives the additional income; when the median household receives the additional income, the Gini decreases slightly because the mean has slightly increased. Furthermore, considering formula (4) as a function of ε routine calculus shows that for j> (n+1)/2 the increase in the numerator has a greater effect than the increase in denominator, so the *largest increase* in the Gini index occurs when the additional increment goes to the top-ranked household.

It is interesting to determine the ranks of the households who can receive an order preserving increment (ε) and still the Gini index decreases. The first step is to calculate the difference between g_1 and the original g. From (2), it follows that

$\sum_{i=1}^{n}(2i-n-1)x_i$ = g$\bar{\boldsymbol{x}}n(n-1)$. Substitution in (4), yields

$$g_1 = \frac{g\,\bar{x}n(n-1)+\varepsilon(2j-n-1)}{(n-1)n\bar{x}+(n-1)\epsilon}. \tag{5}$$

Thus, when the j^{th} ranked household receives the additional small increment, $g_1 < g$ or inequality *decreases* when

2j-n-1<g(n-1) or j<((1+g)n +1-g))/2 (6)

and $g_1 > g$ or inequality *increases* if j>((1+g)n +1-g))/2 .

When n is large, equation (6) implies that inequality will still be reduced as long as the recipient is less than the $(100(1+g)/2)^{th}$ or the $(50 + 50g)^{th}$ percentile. When the original Gini index is .50 (.30), this implies that as long as the recipient is in the lower 75% (65%) of the distribution, inequality as measured by the index will decrease. On the other hand if the small increment goes to a household in the upper 25% (35%), inequality will be increased by an amount that increases with the rank of the recipient. For any value, g, of the original Gini index, as expected the *largest increase* in inequality occurs when the household with the *largest* income receives the small increment and the *greatest decrease* occurs when the household with the *lowest* income receives it. Regardless of the original value of the Gini index, it will decrease as long as a household in the lower half of the distribution receives it; however, the magnitude of the decrease is greatest when households at the lower end receive the increment.

Although the above analysis does not require the order preserving transfer or increment to be small, in practice, the size (n) of the population will be reasonably large so the order preserving requirement restricts the possible magnitude of the transfer or increment.

Thus, the results in this section are applicable to the "infinitesimal" transfers considered by Atkinson (1970) and the "small ones" discussed by Allison (1978). In the order preserving context, neither transfers or an addition to one member of the population in the middle receive excessive weight compared to other parts of the distribution, so the Gini index is *not* especially sensitive to these types of change in the middle part of the distribution.

4. The effect of a transfer that changes the ranks while preserving the mean

Consider the case where the amount, a, transferred by the j^{th} highest income recipient to the i^{th}, where i<j, is sufficient to change the ordering of the n incomes. This means that either x_i +a is larger than some of the observations, x_r, where r>i or x_j – a, is now smaller than some of the observations x_s that previously were below x_j or both occur. In the first case, suppose that x_i +a now is the k^{th} largest. Then the k-i, observations, x_r , r=i+1,…,k that were larger than x_i were *less* than x_i +a. After the transfer, the ranks of each of these k-i observations *decrease* by one, so each of their contributions to the numerator *decreases* by $2x_r$. The contribution of x_i +a, which is now the k^{th} largest observation is (2k-n-1)(x_i +a), however, this replaces its previous value (2i-n-1)x_i, so the transfer increases the contribution of x_i by

(2k-n-1)(x_i +a) - (2i-n-1)x_i = 2(k-i)x_i +a(2k-n-1). Thus, the contribution of the first k observations to the numerator has changed by

$$-\sum_{r=i+1}^{k} 2x_r + 2(k-i)x_i + a(2k-n-1) = -\sum_{r=i+1}^{k} 2(x_r - x_i) + a(2k-n-1). \quad (7)$$

If the rank j of the donor is *unchanged*, its contribution to the numerator of g is reduced by (2j-n-1)a, so the net change in the numerator is:

$$-\sum_{r=i+1}^{k} 2(x_r - x_i) - 2a(j-k)\ . \quad (8)$$

As there are k-i terms in the summand, adding and subtracting 2a(k-i) yields

$$2\sum_{r=i+1}^{k} [a - (x_r - x_i)] - 2a(j-i). \quad (8a)$$

Formula (8a) shows that the effect of a transfer of size a from the j^{th} ranked household to the i^{th} ranked, raising it to rank k, where i<k<j but does *not* change the rank of the donor, depends on the size (a) of the transfer, the difference (j-i) between the before transfer ranks of the donor and recipient and the number (k-i) of households in the interval [x_i, x_k], i.e., in the summand term of (8a). Their incomes satisfy $x_i < x_r < x_i + a$, so $x_r - x_i < a$ implying that each term in the summand is positive but less than a. Thus, the summand in (8a) is less than 2a(k-i), so expression (8a) is *negative* and the Gini index *decreases* whenever (j-i) exceeds (k-i), i.e. the amount transferred is not large enough to reverse the rank order of the recipient and donor. This condition always holds for transfers satisfying the Pigou-Dalton criteria, which must decrease the index of inequality.

Consider the implications of formula (8a) when a high income donor transfers money to a low income recipient. Because the density function of most income distributions is relatively small in the tails of the distribution the number of terms in the summand will be less than the number of terms involved when the recipient is in the central part of the distribution. This means that the difference (j-i) in the ranks of the donor and recipient is much larger than the difference (k-i) in the before and after transfer ranks of the recipient. Thus, the positive summand is much smaller than the absolute value of the second term and (8a) will be substantially negative and the Gini index should decrease noticeably.

The density function of most income distributions has a mode near the median, so transfers to the middle change the ranks of a *larger fraction* of the population than a transfer to the lower part of the distribution. This implies that the number of terms in the summand is larger when the i^{th} household, the recipient is in the middle of the distribution than when it is in the lower portion of the distribution. Thus, the contribution of the summand, which is positive, should be larger when the recipient is near the middle of the distribution than when it is in the lower end. This first term, however, is offset the negative, term -2a(j-i), which reflects the difference between the donor's rank and the recipient's one. This term is smaller when the recipient is in the middle; therefore, the decrease in the numerator of the Gini index as well as the Gini index should be *less* when the recipient is in the central portion than when the recipient is in the lower tail of the distribution. Thus, the Gini index is *not* more sensitive to transfers from a wealthier household to one in the middle of the distribution than it is when the recipient is in the lower portion of the distribution.

When the transfer is between two households in the central portion of the distribution the summand in equation (8a) will increase because the number of households (k-i) who had somewhat higher income than the recipient but now have less will be larger. However, the term 2a(j-i) will be *smaller* than it is when the donor is in the upper income range as the rank j of the donor now is in the middle. Thus, the net effect of a transfer between two middle income households is to decrease the Gini index by a smaller amount than a transfer from a high income donor to a middle income recipient. The Gini index will decrease even more; however, when the recipient is in the lower part of the distribution as the density function is lower than it is in the middle so the lower income recipient will pass over fewer households than when the recipient is in the middle part. This implies that the first term will be smaller in this case than when the recipient is in the middle. Thus, transfers from an upper or middle income household to a middle-income one does *not* affect the Gini index more than a transfer of the same size from the same donor to a low income household. In fact, the index is more sensitive to transfers to the lower end than it is to transfers involving the middle

4b. The transfer increases the rank of the recipient and decreases the rank of the donor.

Next, assume that the size (a) of the transfer from the j^{th} ranked household in the original distribution is sufficiently large that its rank after the transfer will decrease to t, where $t<j$ and the recipient's rank increases from i to k, where $k<t$. This means that the observations x_r where r=t, t+1,…,j-1, are less than x_i but larger than x_i –a. The contributions of each these observations to the numerator will increase by $2x_r$, while contribution of x_j will change by $(2t-n-1)(x_j-a) –(2j-n-1)x_j= -2(j-t)x_j –a(2t-n-1)$. This causes the numerator of the Gini index of the new data to differ from the numerator of the Gini index of the original data by

$$2\sum_{r=t}^{j-1}(x_r - x_j) - a(2t-n-1) = -2\sum_{r=t}^{j-1}(x_j - x_r) - a(2t-n-1) \quad (9)$$

Thus, when the transfer of the amount a from the jth ranked to the ith ranked household increases the rank of the recipient to k, while the donor's rank is reduced to t, it follows from (7) and (9) that the Gini index, g_1 of the new data is:

$$g_1 = \frac{g\,\overline{x}n(n-1) - \sum_{r=i+1}^{k} 2(x_r - x_i) - 2\sum_{r=t}^{j-1}(x_t - x_r) - 2a(t-k)}{\overline{x}n(n-1)}. \quad (10)$$

Provided that the after transfer rank of the donor is larger than that of the recipient the three terms in (10) reflecting the transfer are negative because a higher ranked household is transferring income to a lower ranked one. Recalling that the $x_r - x_i$, r=i+1,…,k and the terms $x_t - x_r$ are less than a, it follows that.

$\sum_{r=i+1}^{k} 2(x_r - x_i)$ <2a(k-i) and $2\sum_{r=t}^{j-1}(x_t - x_r) < 2a(j-t)$.

Hence, the magnitude of the contribution of the first two of the three terms is less than 2a times the number of households not involved in the transfer whose ranks were changed by it. Only when this number, k-i +j-t is comparable to the difference between the ranks of the donor and recipient after the transfer has taken place, will the second and third terms in the numerator of (10) have a major impact on the change in the Gini index. When the donor is in the upper portion of the distribution while the recipient is in the lower part, the term 2a(t-k) indicating the difference in their after-transfer ranks will be much larger than the term reflecting the number (k-i) of households who were passed by the recipient plus the number (j-t) now having larger incomes than the donor. In this situation, the decrease in the Gini index due to the difference in the after-transfer ranks of the donor and recipient will have a greater impact on the decrease in the Gini index due to the number whose ranks were affected by the process.

When the recipient is in the middle of the distribution, e.g. the median, m, there will be more households passed over by the now former median who receives a from the donor (jth ranked), but if the donor is in the upper part of the distribution and a is small, j-t will also be small, the dominant term in the part of the numerator of (10) reflecting the impact of the transfer on the Gini index, remains the difference in the after-transfer ranks of the donor and recipient. When both the recipient and donor are in the middle of the distribution, the higher density of the income distribution in the central region implies that the number of terms in the second and third terms in the numerator of (10) is larger than in the situation where a high income donor makes a transfer to a low income one. The term, involving the difference in the after-transfer ranks of the donor and recipient, however, will be *smaller* than when a high income donor makes a transfer to a recipient in the low or middle part of the distribution. Thus, the net change in the Gini index arising from a transfer between two households in the middle will not be much larger than that resulting from a transfer one involving two households with a large difference in both before and after transfer ranks.

It is useful to consider a simple example when the rank-orders of the population are changed as a consequence of the transfer or addition. For 2013, the U.S. Census Bureau reported the average incomes of the five quintiles as: 12651,30509,52322,83519,184206. To create a data set to explore the effect of changes in the middle, we added four incomes, three in the middle and one in the upper end so that the final set of nine observations has the same mean (72641.40) and median (52322) as the five quintiles. This set is: 11651, 30509, 48322, 50322, 52322, 54322, 83519, 137599.6, and 185206. Their Gini index is .41677 and their mean difference is 60549. Table 4.1 presents a few examples of the change resulting from a transfer of $10000 from a household with a higher income to one with a lower income. This large amount was chosen to illustrate the effect of the transfer increasing the recipients rank or reducing the donor's rank. The transfers between the sixth and fifth as well as the fifth to fourth violate the Pigou-Dalton condition as the recipient's after transfer rank is lower than the recipient's, which is why these transfers increases the Gini index. The last transfer in Table 4.1, from the fourth to fifth is a regressive transfer, which increases the Gini index.

Table 4.1: The effect on the Gini index and mean difference of a transfer of $10,000 from the household j^{th} ranked to the i^{th} ranked household.

Rank of donor	Rank of Recipient	Gini index	Change in the Gini index	Mean Difference
9(highest)	1 (lowest)	.3862	-.0306	56105
9	2	.39	-.0268	56660
9	5	.4045	-.0123	58771
9	8	.4129	-.0039	59994
6	1	.4045	-.0123	58771
6	2	.4084	-.0084	59327
6	5	.4237	.0069	61549
5	4	.4244	.0076	61660
4	5	.4267	.0099	61994

Note: The mean difference is the expected value of the absolute difference of a random pair from a population. It is calculated from a random sample on size n using formula (1).

Examining Table 4.1, the largest change in the Gini index occurs when the richest (rank 9) donates to the poorest (rank 1). As the rank of the recipient of a donation from the richest increases, the magnitude of the change in the Gini index and mean difference decrease as expected. When the sixth ranked household donates to the poorest, the Gini index changes more than when the recipient is the second ranked; however, when the sixth ranked donates to the fifth ranked, the Gini index and mean difference are greater than their values on the original set of nine. The absolute values of the change in the Gini index and mean difference, however, are smaller than when the sixth ranked household transfers money to the lowest ranked. Transfers between the fourth and fifth ranked households also create a slight increase in the Gini index, again due to the lowered rank of the donor, but the magnitude of this change is smaller than the transfers from the ninth ranked household to the poorest or the median one. Again, transfers involving the two households in the central part of the distribution do *not* create the largest changes in the Gini index.

In sum, the change in the Gini index due to a non-order preserving but mean preserving transfer depends on the difference between the after transfer ranks of the donor and recipient and the number of other households whose rank changed. The relative weight of the two components depends on whether the difference in these ranks of the donor and recipient is large or small. When an upper income household makes a transfer to one in the lower or central portions of the distribution, the term involving the difference in ranks is the dominant one, while when the transfer is between two households in the middle; the number of other individuals whose rank changed in the process has a larger role. The fact one component predominates in the change in the Gini in one situation while the other component has more of an effect in another situation does *not* imply that the Gini index is overly sensitive to transfers involving the middle of the distribution. As Atkinson (1970) noted the Gini index is sensitive to changes at all levels and the reason for this is that the change is reflected in two components, one of which is more sensitive to transfers when the donor and recipient are from different parts of the distribution and the other when they are both in the central region.

5. The effect on the Gini index when one household receives a small increment.

Finally, consider the effect of an *additional* amount a being given to the i^{th} ranked household, which increases the mean to $(n\bar{x}+a)/n$. The denominator of g_1 replaces the original mean, $\bar{x}$ by the new mean, is the same regardless of which household receives the additional amount, a. Thus, the sensitivity of the numerator to which member of the population receives the increment is of interest. If the rank of the recipient is unchanged, which is likely when a is small, formula (2) implies that the numerator of the Gini index will change by (2i-n-1)a. If i=1, the numerator changes by –a(n-1) and if i=n, it increases by a(n-1). When n=2m-1 and the median receives the increment the numerator does not change. Indeed, the absolute value of (2i-n-1) declines from n-1 to 0 as j ranges from 1 to m and then increases to n-1 as i ranges from m to n. Clearly, the numerator of the Gini index changes *less* when the recipient is in the middle part of the distribution than when a low or upper income household receives the increment.

If the increment a, is sufficient to effect the order, the rank of the recipient increases by the number (b) of observations between x_i and x_i +a. Then the rank of the recipient increases to i+b=k. The reasoning in the previous section shows that the numerator of the Gini index of the new data differs from that of the original data by

$$-\sum_{r=i+1}^{k} 2(x_r - x_i) + a(2k-n-1). \qquad (11)$$

The $x_r, r = i = 1, \ldots, k$ fall in the interval $[x_i, x_i + a]$ so the first term in (11) is negative. The second term is negative if the final rank, k, of the recipient is less than (n+1)/2 and positive otherwise. The first term is influenced by the number of observations between x_i and x_i +a, which is largest when the rank, i, of the recipient is in the middle of the distribution because the density function is highest there. The second term in (11) tells us that when the final rank, k, of the recipient, is less than (n+1)/2, the numerator will *decrease*. Because the first term in (11) is negative, the Gini index will decrease for some values of k >(n+1)/2, provided they satisfy

$$\sum_{r=i+1}^{k} 2(x_r - x_i) > a(2k-n-1). \qquad (12)$$

For the remaining values of k, all of which are in the upper half of the distribution, the Gini index will *increase*. When the household with the lowest income is the recipient both terms in (11) are negative provided the amount (a) is not so large that their after transfer rank (k) is in the upper half of the distribution and no longer satisfies (12). Thus, for small increments analogous to the transfers considered by both Atkinson (1970) and Allison (1978), the Gini index will decrease the most when the household with the lowest income is the recipient.

When a is small, the first term in (11) is approximately -a(k-i) as the x_r are approximately uniformly distributed in the small interval [x_i , x_i +a] and the change in the numerator of the Gini index is approximately a(k+i) –a(n+1) =a(2i +k-i) –a(n+1). Recall that (k-i) is the number of households the recipient passes, which is largest when the recipient is in the middle of the distribution, because the density is higher there and smallest when the recipient is in the lower or upper parts of the distribution. Therefore, the positive term a(2i+k-i) is smaller when the recipient is in the lower part of the distribution than in the middle, which means that the decrease in the Gini index is larger when the recipient of the small increment is in the lower part of the distribution and that the Gini index is more sensitive to a low income household receiving a small additional income than it is to a middle income household receiving it.

When a household in the middle of the distribution receives the small amount, a, their new rank (k) should also be in the central region. When n=2m-1 and the recipient is the median, the second term in (11) is zero and the decrease in the numerator is due to the number of households passed. This implies that the absolute value of the second term in (11) will be smaller when the recipient is in the middle than when the recipient is in the lower or upper portion. More generally, when a middle income household receives the increment, the first term will have a larger part; however, the decrease due to the first term is ≥ -2a(k-i). Because the effect of the second term in (11) is small in this context, the total *decrease* in the Gini index when a middle income household receives the increment will be *less* than the decrease, occurring when the lowest income household receives it.

For the remaining values of k, which do not satisfy (12) and must be in the upper half of the distribution, the Gini index will *increase*. If the household with the highest income receives the increment, the original order is preserved and the first term in (11) is zero, and the numerator increases by a(n-1), which is *larger* in absolute value than the decrease in the Gini resulting when the median income household receives the small increment.

The conclusion that additional small increments given to households in the lower (upper) part of the distribution decrease (increase) the Gini index is quite intuitive. The result that the effect on the Gini index of an additional increment given to a household in the middle sufficient to change the ordering is *less* than the effect of the same increment given to a household in other parts of the distribution, such as the lower and extreme upper, is less obvious. These changes in the Gini index arising from one household receiving additional income are *not* what would be expected if the Gini index were more sensitive to changes in the middle of the distribution than elsewhere.

To illustrate these results consider our example of nine incomes and assume that an additional $24000 is given to one of them. This value was chosen to ensure that the rank of the recipient would increase in some cases. For each of the nine possible recipients the new values of the Gini index and the resulting change and mean difference are given in Table 5.1. The number of households who had more income originally but the recipient now exceeds is the number passed. Recall that the Gini index, mean difference and mean of the original nine are: .41677, 60549 and $72641.60, respectively. The ratios of the new Gini index to the original one are .8842, 1.0192 and 1.0495 if the first (lowest), median, or ninth (highest) income receives the addition. In percentage terms, the Gini index changes the most when the poorest member of the population is the recipient and the change when the median member receives it is less than the change when the richest member receives it. Thus, the Gini index is *more sensitive* to additional income going to a household in the two extreme regions of the distribution than it is to a similar change in the middle of the distribution.

Table 5.1: The effect on the Gini Index and Mean Difference from one of the nine households receives an additional $24000.

Recipient	Number passed	Gini Index	Change in the Gini Index	Mean Difference
First(lowest)	1	.3685	-.0483	55502
Second	4	.3802	-.0366	57257
Third	3	.4064	-.0104	61216
Fourth	2	.4087	-.0081	61548
Fifth	1	.4101	-.0067	61711
Sixth	0	.4109	-.0059	61833
Seventh	0	.4197	.0029	63216
Eight	0	.4286	.0118	64549
Ninth(highest)	0	.4374	.0206	65883

6. Summary and Discussion

Although the Gini index yields one number summarizing the entire income distribution or Lorenz curve and cannot capture all the changes in the income distribution that economists or policy makers are interested in, the scenarios studied here indicate that the criticism that it gives undue weight to changes in the middle of the distribution is inaccurate. In particular, when a transfer or increment preserves the order, e.g. when it is small, the *opposite* is true. In most situations the difference in the ranks of the donor and recipient, either before or after the transfer or addition is the main contributor to a change in the Gini index. Transfers or an additional increment involving a middle income household do change the ranks of a *higher fraction* of the population as a consequence of the density function being higher in that region. Only when the transfer is between two households in the middle of the distribution did the number of individuals whose rank was affected by the transfer have an important role. This effect on the change in the Gini index is offset by the relatively small difference between the after transfer ranks of the donor and recipient when both are in the central region. Thus, small transfers or a small additional increment affecting the middle of the income distribution do not have undue weight on the resulting change in the Gini index. Indeed, in the scenarios examined here a transfer or addition to a middle income household had a smaller impact on the Gini index than a transfer or addition to a low income household. In view of the larger magnitudes of the weights the numerator of the Gini index in formula (2) assigns to the more extreme ordered incomes than it does to the middle ones, in the context of small transfers or additions, these conclusions are not that surprising.

When the total income of the same population and consequently the mean increase; because twice the mean is the denominator of the Gini index, as illustrated by Table 5.1, it does not fully reflect a shift in favour of the upper end. Indices such as the ratio of the share of income received by the top 20% to the lower 20%, used by Dorling (2014), the ratio of the share of the top 10% to the lower 40% introduced by Palma (2011) or the median based Gini index (Gastwirth, 2014) increase more than the Gini index in response to such a change. Like the Gini index, however, these indices can have the same numerical value for data from two distributions even though the Lorenz curves intersect. A method for constructing two different with same value of a measure that is the ratio of the top 100u% to the bottom 100b%, where $b< 1-u$, is described in Appendix A.

From a statistical viewpoint it is unreasonable to expect one summary measure to capture the features of an entire distribution. Thus, the relationship between the choice of measure and its underlying social welfare function, stressed by Atkinson (1970), Newberry (1970), Sheshinski (1972) and Sen (1974) remains very important. Jenkins (2009) noted that the ability to calculate several indices, which focus on changes in different income ranges, is very useful for the analysis of income and earnings distributions. By using different weights than those in the numerator of formula (2), following Mehran (1976) one can create a summary measuring placing increased weight on the part of the income distribution most relevant to the purpose of a study. Like the numerator of the Gini index, these measures are linear combinations of order statistics and there is a large literature deriving their large sample distributions (see Greselin et al. 2009).Similarly, the Lorenz curve and related functions or transformations of it (Sordo et. al. 2013, Arnold, 2015; Gastwirth, 2016) can be used to emphasize the most appropriate region of the income distribution.

Although other measures of inequality, e.g. the generalized entropy family (Cowell, 2011) or Atkinson's family may be superior to the Gini index for some analytic purposes, it may be difficult for an agency producing income statistics to choose one member of the family, which would place greater emphasis on one part of the distribution. The Gini index is a well-studied index with a long history (Giorgi, 1990) and is associated with the area between the line of equality and the Lorenz curve; providing a graphical summary of the distribution that economists and policy makers have found useful. Government agencies should supplement the Gini index with information on the mean, median, deciles and 95th and 99th percentiles of the distribution and the shares of each decile and the top 5 and 1 percent. If statistical agencies would provide a sufficiently detailed grouping of the data, researchers would be able to accurately estimate the measure of inequality appropriate for their project and preserve the confidentiality of the incomes of the survey respondents.

Acknowledgements: The author would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme *Probability and Statistics in Forensic Science* which was supported by EPSRC grant number EP/K032208/1. It is also a pleasure to thank Dr. Barry Graubard and Professors Weiwen Miao, Julia Mortera and Yulia Gel for helpful discussions.

Appendix A: An example of two crossing Lorenz curves with the same value of the Palma Index.

The Palma (2011) index is the ratio of the share of income received by the top ten percent to the share of income received by the lower forty percent. Because it focuses on the status of the lower part of the distribution relative to the upper end, it is more responsive to changes in the distribution that favour the upper end and disadvantage the lower end than the Gin index. Like the Gini index, however, two income distributions can have intersecting Lorenz curves and the same value of the Palma index.

To construct the example, begin with the Lorenz curve $L_1(p)=p^2$, corresponding to a uniform distribution on the unit interval. The Palma index = (1-L(.9))/L(.4) = 1.1875. The example will have the same values $L_1(.9)=.81$ and $L(.4)=.16$ as $L_1(p) = p^2$ but will be less than p^2 in the interval [.4, .5] and then increase faster to reach .81 at p=.9. Recalling the lower bound for the Lorenz curve in Gastwirth (1972), the slope of the tangent line to $L_1(p)$ at p=.4 is 2p=.8. Hence, the line $T_1(p)=.16 +.8(p-.16)=.16+.8p$ equals $L_1(p)$ at p=.4 and is below it on [.4,.5], and equals .24 when p=.5. The line $T_2(p)$ connecting the points (.5,.24) to (.9, .81) is .24 + .57(p-.5)/.4 or .24 +1.425(p-.5) over the interval [.5,.9] must cross the original Lorenz curve $L_1(p)$. Indeed, one can verify that $T_2(.525) =.525^2= 0.275625$. Hence, the Lorenz curve defined by $L_1(p)$ over the intervals [0,.4], $T_1(p)$ over [.4,.5], $T_2(p)$ over [.5,.9] and $L_1(p)$ on [.9,1] has the same value of the Palma index as $L_1(p)$, even though the two Lorenz curves cross.

The same procedure can be used to construct an example of two intersecting Lorenz curves with the same ratio of the share of the top 20% to bottom 20% or any similar ratio. Furthermore, the method used to construct the crossing Lorenz curves does not depend on starting out with the Lorenz curve for the uniform distribution. One can begin with any Lorenz curve L(p) and construct the tangent line at p=.4; choose a point in [.4,.9] and connect it to L(.9). The only restriction is that the slope of the connecting line must be less than the slope of L(p) at p=.9, in order to preserve the convexity of the Lorenz curve. In our example, that slope 2p=1.8 which is larger than 1.425, the slope of $T_2(p)$.

References

Ahn, K (1997) Trends in and Determinants of Income in Korea, *Journal of Economic Development*, 22, 27-56.

Alison, P.D. (1978). Measures of Inequality *American Sociological Review*, 43, 865-880.

Alison, P.D. (1979). Reply to Jasso. *American Sociological Review*, 44, 870-872.

Atkinson, T.A. (1970), On the Measurement of Inequality. *Journal of Economic Theory*, 2, 244-263.

Arnold, B.C. (2015), On Zenga and Bonferroni Curves, *Metron*, 73, 25- 30.

Australian Bureau of Statistics (2015), Survey of Income and Housing, User Guide, 2013-2014.

Bird, R.M. and Zolt, E.M. (2015). Taxation and Inequality in Canada and the United States: Two Stories or One? *Osgoode Hall Law Journal*, 52, 401-426.

Borghi, E. (2005) Trade openness and wage distribution in Chile. Report 173, Centro di Recerca sui processi di Innovazione e Internazionalizzione.

Chang, H-J.(2014). *Economics: The User's Guide*, London: Penguin Books.

Cobham, A. and Sumner,A. (2013). Putting the Gini back in the bottle? 'The Palma' as a policy-relevant measure of inequality. Kings College London, March 15, 2013.

Cobham, A., Schlogl, L. and Sumner, A. (2015). Inequality and the Tails: The Palma proposition and ratio revisited. DESA Working Paper No. 143. Kings College, London.

Cowell, F. (2011), *Measuring Inequality* (3rd ed.) Oxford: Oxford University Press

David, H.A. (1968). The Gini mean difference rediscovered. *Biometrika*, 55, 573-575.

David, H.A. and Nagaraja, H.N. (2003) Order Statistics (3^{rd} ed.), New York: John Wiley.

De Maio, F. G. (2007), Income Inequality Measures, *J. Epidemiology and Community Health*, 61, 849-852.

Dorling, D. A. (2014). *Inequality and the 1%*, London: Verso Books.

Fremeaux, N. and Picketty, T. (2014).France: How taxation can increase inequality in *Changing Inequalities & Social Impacts in Rich Countries* ed. by Nolan, B., Salverda, W., Checci, D., Marx, I, McKnight, A., Toth, I.G. and van der Werfhorst, H. Oxford: Oxford University Press.

Gastwirth, J.L. (1972), Robust estimation of the Lorenz Curve and Gini index. *Review of Economics and Statistics*, 54, 306-316.

Gastwirth, J.L. (2014). Median-based measures of inequality: Reassessing the increase in income inequality in the U.S. and Sweden. *Statistical Journal of the IAOS*, 30, 311-320.

Gastwirth, J.L. (2016). Measures of Economic Inequality Focusing on the Status of the Lower and Middle Income Groups. *Statistics and Public Policy*, 3, 1, 1-9.

Giorgi, G.M. (1990). Bibliographic portrait of the Gini concentration ratio, *Metron* 48, (1990), 183–221.

Green, G.W, Jr., Coder, J. and Ryscavage, P. (1994) Earnings inequality for men in the 1980's in *Aspects of Distribution of Wealth and Income* ed. by P. Papadimitriou. London: Macmillan.

Greselin, F., Puri, M.L. and Zitikis, R.(2009), "L-functions, Process and Statistics in Measuring Economic Inequality and Actuarial Risks," *Statistics and Its Interface*, 2 (2), 227-246.

Jasso, G. (1979). On Gini's mean difference and Gini's index of concentration, *American Sociological Review*, 44, 870-872.

Jenkins, S.P. (2009), Distributionally-sensitive inequality indices and the GB2 income distribution. *Review of Income and Wealth*, 55, 392-398.

Jones, A.F. Jr. and Weinberg, D.H. (2000). The changing shape of the nation's income distribution. Current Population Reports P60-204. U.S. Census Bureau: Washington DC.

Kendall, M.G. and Stuart, A. (1977), *The Advanced Theory of Statistics*, Vol. I. London: Griffin.

Madden, J.F. (2000) *Change in Income Inequality within U.S. Metropolitan Areas*. Kalamazoo, MI: W.E. Upjohn Institute.

Mehran, F. (1976), "Linear Measures of Income Inequality," *Econometrica*, 44, 805-809.

OECD (2013). OECD Guidelines for Micro Statistics on Household Wealth.

Palma, G. (2011). Homogeneous Middles vs. Heterogeneous Tails, and the end of the "Inverted-U". *Development and Change*, 42, 87-153.

Pressman, S. (2013).Cross National Comparisons of Poverty and Income Inequality, in The Economics of Inequality, Poverty, and Discrimination in the 21st Century ed. Rycroft, R.S., Santa Barbara, CA: Praeger.

Roberts, A. and Willits, D. (2015). Income inequality and homicide in the United States: Consistency across different income inequality measures and disaggregated homicide types. *Homicide Studies*, 19, 28-57.

Schmid, K.D. and Stein, U.(2013). Explaining rising income inequality in Germany: 1991-2010. Dusseldorf: Macroeconomic Policy Institute (IMK) Study 32.

Sen, A. (1974). Informational bases of alternative welfare approaches. *Journal of Public Economics*, 3, 387-403.

Sordo, M.A., Navarro, J. and Sarabia, J.M. (2014), “Distorted Lorenz curves: models and comparisons,” *Social Choice and Welfare*, 42, 761-780.

Sudheesh, K.K. and Dewan, I. (2013). A simple non-parametric estimator of the Gini index. Preprint series, Indian Statistical Institute at Delhi.

Thewissen, S., Kenworthy, L., Nolan, B., Roser, M. and Smeeding, T. (2015) Rising Income Inequality and Living Standards in OECD Countries: How Does the Middle Fare? LIS Working Paper Series, No. 656.

Thon, D.(1982), An Axiomatization of the Gini Coefficient. *Mathematical Social Sciences*, 2, 131-143.

Yitzhaki, S. and Schectman, E. (2013). *The Gini Methodology: A Statistical Primer*, New York: Springer.

Yitzhaki, S. (2003). Gini’s mean difference: a superior measure of variability for non-normal distributions. *METRON*, 61, 285-316.