A Summary measures of the dispersion of income

A wide variety of methods can be used to illustrate and measure the degree of income dispersion in a population.

Incomes vary between individuals and households for a number of reasons which are often complex — for example, workforce attachment (full- or part-time work) and wage rates are driven by a number of factors including educational qualifications, experience and employment opportunity. Summary measures of inequality alone, therefore, provide very limited information about the distribution of income. They should be interpreted as statistical indicators of the dispersion of income only, rather than being used as indicators of underlying causes of differences or other wider interpretations.

This appendix describes the most commonly used methods and inequality indexes which have been used in this study and briefly describes their relative strengths and weaknesses. The measures discussed include:

* percentile ratios and income shares
* the Lorenz curve
* the Gini coefficient
* the generalised entropy indices.

All of these measures are scale invariant — that is, a uniform increase in income for all does not change the measure of dispersion. Usual measures of variance (or standard deviation) are scale dependant making them inappropriate for measuring the distribution of income across population groups (Jenkins and Van Kerm 2008).

## A.1 Percentile ratios and income shares

Income shares and percentile ratios are the simplest measures of income dispersion. Commonly used indices include the ratio of income levels of the 90th to 10th percentile (‘P90:P10 ratio’), and as a comparison of dispersion at the top and bottom of the distribution, the P90:P50 ratio and the P50:P10 ratio.

The income estimate at a given *x* percentile represents the value of income below which *x* per cent of all incomes fall. For example, the income value at the 50th percentile means that 50 per cent of people earn less than that income (the 50th percentile is the same as the median income). Income at the 90th percentile means that 90 per cent of the population earn less than that amount.

One of the advantages put forward for using percentile ratios is that they avoid the problem of ‘top-coding’ in survey data (where instead of the actual income being reported, it is only reported as greater than some amount). They can also be used to determine the ‘location’ of dispersion — that is, among which groups income disparities are the greatest. However, percentile ratios focus on the percentiles examined and therefore ignore information about the spread of incomes that occur elsewhere in the distribution.

A related measure is the examination of income shares. Income shares describe the share of total income accruing to particular groups. It is most often used to describe how concentrated income is at the top of the income distribution, in particular the highest 10 per cent, one per cent and 0.1 per cent of earners. These ratios are constructed by expressing the total income earned by those in, for example the top 10 per cent, as a share of total income earned by the entire population (or sub-group). Again, as with percentile ratios such measures can ignore information on the spread of incomes that occurs elsewhere in the distribution.

## A.2 Lorenz curve

Income ratios are only able to compare two parts of a distribution rather than describe the distribution as a whole. An alternative to ratios is the depiction of the spread of incomes within the population through the Lorenz curve (and indices derived from it such as the Gini coefficient — see following section).

The Lorenz curve, named after Max Otto Lorenz (1905), maps the cumulative proportion of the population, ranked by income, against their cumulated share of income (figure A.1). If all individuals within a population earn the same income the Lorenz curve is a 45 degree line (often termed ‘perfect equality’). When individuals within the population earn differing amounts, the curve takes on an exponential shape and lies below the 45 degree line (curves A, B and C in figure A.1). The larger the gap between the Lorenz curve and the 45 degree line, the less equal the income distribution. That is, Lorenz curve A represents a more equal income distribution than curves B and C. If the Lorenz curves of two populations cross each other, such as Lorenz curves B and C, it is difficult to determine in which population income is distributed more equally. However it is still possible to identify differences in the distributions quickly, for example the lowest income earners in population C earn a larger share of total income than the lowest earners in population B.

Figure A.1 Lorenz curve

Per cent

|  |
| --- |
| Depicts various Lorenz curves. More details can be found within the text immediately before this image. |

## A.3 Gini coefficient

The Gini coefficient, named after the Italian statistician Corrado Gini, is a popular summary statistic for measuring the dispersion of income across a population. It is defined as the ratio of the area between the 45 degree line and the Lorenz curve (area ‘a’ in figure A.2) and the area under the 45 degree line (areas ‘a’ + ‘b’ in figure A.2). If individuals with a population all earn the same amount the Lorenz curve lies on the 45 degree line and area ‘a’ is equal to zero, hence the Gini coefficient is equal to zero. At the other extreme, if one person in a population earns all the income, the Lorenz curve runs along the x-axis and up the right hand side of the diagram meaning area ‘b’ is equal to zero and the Gini coefficient is equal to one.

In reality, populations will lie somewhere between these extremes. Populations with more equally distributed income have a Gini coefficient closer to zero, those with less equally distributed incomes have a Gini coefficient closer to one.

Figure A.2 Lorenz curve and the Gini coefficient

Per cent

|  |
| --- |
| This figure depicts the calculation of the Gini coefficient graphically using the Lorenz curve. The Gini coefficient value is the ratio of the area under the Lorenz curve (which depicts the cumulative income shares against cumulative population share) to the area under a 45 degree ray from the origin. |

There are numerous equivalent methods of calculating the Gini coefficient, two examples are shown below. The left-hand side equation is derived from calculating the area under the Lorenz curve for a population and using this to calculate the Gini coefficient. The right-hand side equation is calculated from the covariance between a population member’s income level (such as a household or individual) and their cumulative rank amongst all members of the population.

,   
*n* is total number of individuals*, Yi* denotes disposable income of household *i,   
Yij* denotes the adjusted equivalised income of each member*, j,* of household *i,* and household incomes per equivalent household members *(Yij=Yk)* are ranked in ascending order (such as *k = 1, 2, ....n*)*.*

While indices such as the Gini coefficient make use of information across the entire income distribution (unlike percentile ratios and income shares), they also have a number of shortcomings. For example, two societies with the same Gini coefficients can have very different distributions of income. Consider a hypothetical country (country A) in which 50 per cent of the population receives all the income earned in equal amounts (that is, total income is distribute equally amongst 50 per cent of the population (figure A.3, left panel). The Gini coefficient for country A would be 0.5. Another country (country B) where 25 per cent of all income is earned equally by 75 per cent of the population with the remaining income earned equally by 25 per cent of the population would also have the same Gini coefficient — 0.5 (figure A.3, right panel). Despite having the same Gini coefficient, these countries have very different underlying income distributions as seen when examining the Lorenz curves (figure A.3).

Figure A.3 Lorenz curves showing different income distributions for the same Gini coefficient

Per cent

|  |  |
| --- | --- |
| Country A Left panel depicts country A. More detail can be found within the text immediately before this image. | Country B Right panel depicts country B. More detail can be found within the text immediately before this image. |

The Gini coefficient, as with other indices, can also be affected by the level of granularity of income data. It has been found that for the same population, higher granularity reduces measured inequality (Monfort 2008). For example, a Gini coefficient estimated using quantile data (low granularity) will usually yield a lower estimate than one based on deciles or quintiles (higher granularity).

## A.4 Generalised Entropy indices

Generalised entropy indices are a family of indices describing income distribution in a population. They are similar to the Gini coefficient in that they produce a single number describing the level of distribution. However, the indices can be adjusted to give greater weight to particular sections of the income distribution (such as weighting disparities between the bottom and top earners more heavily than disparities between earners in the middle of the income distribution). The indices can range from zero, all receive the same income, to infinity, with higher numbers representing wider distributions.

The general formula for Generalized Entropy indices is:

where the value of determines the sensitivity of the index to certain parts of the income distribution (Jenkins and Van Kerm 2008). The larger (more positive) is the more weight is placed on distribution at the higher income end of the spectrum. The smaller or more negative is, the more weight is given to distribution at the low income end of the income spectrum. Two of the most famous members of the generalised entropy index family are the mean logarithmic deviation, where = 0, and the Theil index, where = 1.

Generalised entropy indices have the added advantage of being decomposable into sub-groups, such as age or labour force status (unlike the Gini coefficient which is only fully decomposable if the incomes of the sub-groups do not overlap). The entropy index for the population as a whole is then the weighted sum of the entropy indices for each of the population sub groups. This enables analysis of sub-group effects on the main drivers of the overall income distribution.