San Jose State University

SJSU ScholarWorks

The Measurement of Inequality, Concentration, and Diversification.

Fred Foldvary
Santa Clara University, ffoldvary@gmail.com

Recommended Citation

Fred Foldvary. "The Measurement of Inequality, Concentration, and Diversification." *Journal of Interesting Economics* (2001).

The Measurement of Inequality, Concentration and Diversification

by

Fred E. Foldvary

Department of Economics, Santa Clara University, CA 95053
ffoldvary@scu.edu

Abstract

The Lorenz curve and Gini coefficient are typically used to measure inequality. A different way to measure inequality is introduced here: $I = CN$, the product of concentration and number of units. The resultant index can be interpreted with reference to an inequality base where one unit owns all and the rest nothing. This inequality index also integrates the measurement of inequality, concentration, and diversification into one system, where diversification is measured as the inverse of concentration. $I = CN$ accommodates various measures of concentration, including the Herfindahl-Hirschman and Tideman-Hall indexes. The Tideman-Hall concentration index also provides indexes of concentration, diversification and inequality as functions of Gini. As one application, the inequality index can be used to provide an index of economic development.

1. The meaning of inequality

Inequality can be ordinal or cardinal. An ordinal ranking of distributions in order of inequality implies some cardinal method of measuring or judging each distribution in relation to the others. A cardinal ranking can be relational, defined with reference to other variables, or monadic, its magnitude being defined without reference to other variables (Wright, 1987). This paper is concerned with the measurement of monadic, cardinal inequality.

Inequality is the degree to which the units of a distribution have shares of some attribute which are unequal in quantity. This meaning of inequality is completely general for any type of distribution. In order to measure inequality, one needs to determine the meaning of the degree of being unequal.

Let S be a set of N elements (also called units or members). There is some attribute A (such as income) of which the elements have relative shares a_i, each a_i being a fraction of A. A distribution is designated by $(a_1, a_2, ... , a_N)$ and by $(j(a_i))$, which designates j units of share a_i each. The degree of inequality among the a_i is a function of the concentration of A among a_i and of the number of elements, N. The greater the concentration of a_i, given some N, the more unequal is the distribution. The greater N is, given some concentration, the more unequal the distribution.

Berrebi and Silber (1985) note that inequality indices can be expressed as a income-weighted (more generally, attribute-weighted) sum of individual "deprivation" coefficients, the differences among the indices depending on the way the deprivation is defined. It has been recognized since Dalton (1920) that a measurement of income or wealth inequality implies some concept of social welfare.

It would seem reasonable to posit that, given N, if the concentration doubles, then inequality is also doubled, hence inequality is linearly proportionate to concentration. It would also seem reasonable to posit that, given some concentration C, if N doubles, then inequality also doubles, so that inequality is linearly proportionate to the population size. Putting both variables together, inequality is the product of concentration and number:

$$(1) \quad I = C \cdot N$$

The concentration C of a distribution is often measured by the Herfindahl-Hirschman (HH) index (an alternative, the Tideman-Hall index (12), is discussed below),

$$(2) \quad C_H = \sum_{i=1}^{N} a_i^2$$

Applying (1), the inequality index specific to the Herfindahl-Hirschman concentration index would then be the Herfindahl-Hirshman-Foldvary inequality index,

$$(3) \quad I_H = N \sum_{i=1}^{N} a_i^2$$

The HHF index has the desired property that a transfer of a share from a richer to a poorer member reduces the value of I_H, thus reducing inequality.

Hall and Tideman (1967, p. 165) state that with HH, "the relative sizes of firms are more important than the absolute number of firms in determining concentration in an industry." The index weighs each firm by its relative share. It should be noted, however, that given some accepted inequality index I, that N is as important in computing C as I, so the problem, if any, in C_H would lie in how it takes into account the inequality of a distribution, the weight increasing perhaps too much with size, rather than any neglect of the number of units.

As an example of the effect of changing concentration, consider the distribution (.5, .5), for which C_H = .5. If the distribution is changed to (.4, .6), C_H = .52. The distribution is less equal and has a higher concentration index. For (.3, .7), (.2, .8), and (.1, .9), C_H = .58, .68, and .82. It is evident that inequality increases with concentration. At the extreme, the distribution of "monopossession" (1, 0) has C_H = 1, twice the concentration of (.5, .5). Applying equation (1), the inequality index I_H would equal 1 for (.5, .5) and 2 for (1, 0). I_H always equals 1 for complete equality and N for complete inequality among N elements. The doubling of I_H has an intuitive interpretation: as the share of the first unit doubles from .5 to 1, the inequality also doubles.

For an example of the effect of population size, again consider the "monopossessive" distribution (1, 0), where N = 2 and one member has all of A (i.e. monopolizes income or wealth) and the other has nothing. If the number of elements increases to 4 with the distribution (1, 0, 0, 0), C_H is unchanged at 1. However, applying (1), I_H doubles from 2 to 4. The second distribution is more unequal, though equally concentrated, because in (1, 0), 1/2 the population owns all, while in (1, 0, 0, 0), 1/4 of the population owns all. With complete equality, each would own .25, hence when one member owns all, he has four times the equal share, whereas with (.5, .5), he only has twice the equal share. The inequality index I_H is thus the multiple over pure equality when 1 member owns all of A.

Let I_F designate the class of inequality indexes based on (1), with the concentration index having a range (1/N, 1), 1 being the case of monopossession. For exactly equal distributions, $I_F = 1$. I_F for any distribution can be interpreted as the equivalent in inequality as that of a monopossessive distribution with $N = I_F$. For example, if $I_F = 3$, then the inequality is the same as that with $N = 3$ and one member owning everything. The index I_F thus has a meaning relative to the benchmark of a monopossessive-equivalent distribution. This is consistent with the criterion set by Dalton (1920, p. 349), who stated that "the inequality of any given distribution may conveniently be defined as the ratio of the total economic welfare attainable under an equal distribution to the total economic welfare attained under the given distribution. This ratio is equal to unity for an equal distribution and is greater than unity for all unequal distributions."

Equation (1) thus encompasses both the concentration and population aspects of inequality by multiplying them together, since each by itself increases inequality proportionately. It not only measures inequality but implements the very meaning of inequality in its calculation.

2. Putting the Gini back in the bottle

A common way that inequality is measured is with a "size distribution," in which X percent of the population has Y percent of the attribute such as income. The population and attribute are typically divided into quintiles or deciles. One can also measure the inequality by choosing a ratio between two points in the distribution, such as the percentage of income received by the bottom 10% of the population by the percentage received by the top 10%. But these measurements only provide information about points in the distribution rather than reflecting the entire distribution.

The Lorenz curve, devised by the American statistician Conrad Lorenz in 1905, is widely used to show the relationship between population and shares of income. The cumulative percentage of population is drawn on the horizontal axis, and the cumulative share of income received is plotted on the vertical axis. The diagonal line represents strict equality, with inequality reflected by the amount by which the Lorenz curve deviates from it. Although useful as a visual representation of the inequality of a distribution, it is useful also to have a number that captures the degree of inequality, the Gini coefficient.

Formulated by the Italian statistician Corrado Gini in 1912, the coefficient is the ratio of the area between a Lorenz curve and the 45-degree line to the area of the triangle below the 45-degree line. Its formula is

(4) $$G = 1 + 1/N - (2/(N^2 \bar{a}))(a_1 + 2a_2 + ... + Na_N)$$

where a_i is the amount owned by each member in decreasing order of size and ~ is the mean of the a_i. The Gini index is thus a weighted sum of the shares, with the weights determined by rank order position. As noted by Maasoumi (1995), Gini does not provide for aggregation consistency or full additive decomposability. Also, Gini places more weight to transfers affecting the middle of a distribution than at the tails. However, a function of Gini, such as (14) below, corrects this latter feature.

As half the relative mean difference,

(5) $$G = (1/2)n^2\mu = \sum_{i=1}^{n}\sum_{j=1}^{n}(y_i - y_j)$$

G has a maximum value of unity with extreme inequality and a minimum of zero at

pure equality. However, for distributions with a small N, G is less than one. For example, with (1, 0), G = .5, representing half the triangle. Since the axes represents relative proportions of population, the distribution (.5, .5, 0, 0) also has G = .5, and I_H index for (.5, .5, 0, 0) is 2 as it is with (1, 0), so I_H and G agree with respect to relative inequality where the population increases but there is equality among the "haves".

However, G does not increase proportionate to number when concentration is held constant. For the distribution (1, 0, 0, 0), G = .75, an increase of 50% from (1, 0), in contrast with the doubling of I_H and a decrease of (1-G) of 50%. The monopossessive distribution (1, 7(0)) yields G = .875 and I_H = 8, an increase of 1/6 over (1, 0, 0, 0), while (1-G) again decreases 50%. As N doubles, G decreases in (1-G) by 50% while I_H doubles. At an extreme, a change from (1, 999(0)) to (1, 99999(0)) yields a change in I_H from 1000 to 100,000 while G changes from .999 to .99999, with (1-G) reduced one thousandfold from .001 to .000001. Since G is varying directly with (1-G) rather than with N, G measures little change in inequality if a large N increases while C is constant, while I_F increases proportional to population. With inequality defined as a function of both concentration and population, raw G does not fit (1).

With respect to changes in concentration, the distribution (4(.25)) has C_H = .25, I_H = 1, and G = 0. If we change it to (2(.5),2(0)), then C_H = .5, I_H = 2, and G = .5. For (1, 3(0)), C_H = 1, I_H = 4, G = .75. Here again, I_H has doubled with each doubling in concentration, but G is decreasing (1-G) by half each time, hence increasing proportionately less with increasing concentration, given some population size.

These traits of G are overcome with an inverse-reverse function 1/(1-G) as in the Tideman-Hall-Foldvary inequality index I_T presented below in (14), which yields I_F characteristics for I = CN similar to those of I_H.

3. Inequality, Concentration, and Diversification

Equation (1) not only measures inequality but demonstrates the relationship between inequality, population size, and concentration. Since I = CN, we can also formulate C as

(7) C = I / N

so that concentration equals inequality divided by population. Given some N, the more unequal the distribution, the more concentrated, and given some I, the greater the population, the less concentrated the distribution. Also,

(8) N = I / C

with population size able to be calculated if the inequality and concentration of a distribution are compatibly measured.

Equation (7) provides a general measurement of concentration, given some measurement of inequality, for any type of distribution. The Herfindahl index and the Tideman-Hall index take both N and I into account with weighted sums of the units.

The diversification of a distribution of items, such as the products of a firm, reflects both the number of items and the relative amount of each item. The diversification of an economy reflects the number of industries and relative size of each, and the diversification of assets in a portfolio reflects the number and relative size of the assets. Diversification is thus similar to concentration in being a function of the number and inequality of items, except that as concentration increases, diversification decreases; diversification is an increasing function

of number and decreasing function of inequality. Diversification (D) is thus the inverse of concentration:

(9) $D = 1 / C$

The index D is equal to N when the units are all equal, since

(10) $D = N / I$

diversification being defined and calculated as the number of items divided by their inequality.

The inequality index $I = CN$ hence also provides a method for measuring diversification. Diversification, concentration, and inequality are thus measured by one common system, e.g. using the Herfindahl-Tideman-Hall index for C.

The first three criteria specified by Hall and Tideman (1967) for concentration also apply to inequality. The index I_H satisfies the first Hall and Tideman criterion as an unambiguous one-dimensional measure. It satisfies their second criterion as being a function of all the elements of a distribution and their relative shares. The third criteria is that a change in relative shares changes the index, which is satisfied. Since C_H satisfies all their criteria for C, I using C_H satisfies those that are relevant to inequality.

4. Concentration based on Gini

Suppose that one uses the Gini coefficient G for inequality. At pure equality, $G = 0$, and at pure inequality, $G = 1$. Then by (1), $G = C_G N$, and we have a Gini-based index of concentration:

(11) $C_G = G / N$

With pure equality, $C_G = 0$ regardless of N, hence the application of Gini to concentration provides for a deficient index of concentration. With less than pure equality, C_G will vary with N but still be small for very low G. C_G varies from zero to unity, as does H, satisfying the Hall and Tideman (1967) sixth criterion for the range of concentration.

For diversification, (10) using G provides $D_G = N / G$, and when G is zero, the diversification is infinite. Also, with a very high degree of inequality, G is close to 1 and D_G almost equals N, so D_G would increase with N even if one element monopolizes the distribution. Using I_F, a very high degree of inequality would yield a high I, which would provide a low degree of diversification unless N were correspondingly very high. With G in the range of zero to one, and I_F ranging from one to infinite, I_F provides a more general and meaningful basis for computing diversification and concentration than D_G. I_F thus provides for greater consistency in the measurement of C and D than does G. However, this deficiency in G and its derivatives can be remedied with a function of G.

The Tideman-Hall (TH) concentration index (Hall and Tideman, 1967, p. 166), designated here by C_T, is

(12) $C_T = 1 / (2 \sum i a_i - 1)$

the *i*th largest firm receiving weight i, thus weighing each share by its rank rather than its relative share. Like C , C = 1/N when the distribution consists of N equal units, so I = 1

for an equal equality with C_T as with C_H. For (.9, .1), C_T = .83 versus .82 for C_H, and for (.6, .4), C_T = .56 versus C_H = .52, so both measures are fairly close, and C_T can serve as well as if not better than C_H to provide a Tideman-Hall-Foldvary index I_T of inequality:

$$(13) \quad I_T = C_T N = N / (2\sum_i ia_i - 1)$$

With TH and Gini both rank-weighted, I_T is a function of G:

$$(14) \quad I_T = 1 / (1-G)$$

Hence, with C = I / N,

$$(15) \quad C_T = 1/((1-G)N)$$

I_T as a function of G, with qualities similar to I_H, is in the class of indexes I_F, with a range of 1 for equality to N for maximum inequality. A diversification index using C_T is

$$(16) \quad D_T = 1/C_T = N/I_T = N(1-G)$$

5. Using other measures of inequality

If other measurements of concentration are used besides C_H and C_T, equations (1) and (9) may be used to create the corresponding inequality and diversification indexes. Equations (7) and (10) can also be applied to formulae for inequality, so long as inequality increases with increasing values of the index.

For example, the standard deviation of logarithms measures inequality, taking differences from the mean:

$$(17) \quad I_S = [\sum_{i=1}^{n}(\log\mu - \log y_i)^2 / n]^{1/2}$$

The Theil (1967) entropy-related index is:

$$(18) \quad I_E = \sum_{i=1}^{n} a_i \log(n \cdot a_i)^2$$

where a_i is the fraction of attribute A as above, the entropy of a distribution being

$$(19) \quad E(A) = \sum_{i=1}^{n} a_i \log(1 / a_i)$$

which provides a measurement of diversification.[1]

An equal distribution yields I_E = 0, while a[0] monopossessive distribution, with complete inequality, yields $I_E = \log N$. As Sen (1972) notes, the Theil index is rather arbitrary, but, as Shalit and Yitzhaki (1984) observe, there is an interesting similarity between inequality and decision-making under uncertainty.

Shalit (1985) notes that Gini's mean difference can be used as a measure of dispersion:

$$Ds = 1/2 \int_a^b \int_a^b |x - y| f(x) f(y) dx dy$$

(20)

The relationship between diversification and inequality formulated in (10) provides an inverse relationship, with $I = N / D$ a measure of inequality if diversification is independently measured.

Atkinson (1970) proposes a inequality/welfare measure that is invariant with respect to linear transformations, an "equally distributed equivalent income" that would give the same level of social welfare as the actual distribution:

(21) $I_A = 1 - Y_e / \mu$

The index proposed by Berrebi and Silber (1985) is:

(22) $I_B = \sum_{j=1}^{n} S_j [(n - j) / n - (j - 1) / n]$

where S_j is the share of the j'th rank, $S_1 \$ S_2 \$ ... \$ S_n$.

These various inequality measures, along with the many others that have been proposed, yield different inequality rankings, and, applying (7), yield diverse measurements of concentration. The literature on measurements of inequality thus also provides ways of measuring concentration and diversification.

6. Application to measuring economic development

Michael Todaro (1994) has noted that the concept of economic development has come to be defined as a function of inequality as well as of per-capita GDP. If two countries A and B have an equal per-capita GDP, but A has a much more unequal distribution of income, due to a larger less-developed sector, then it can be regarded as overall less developed than B. Hence, an increase in GDP in a small sector of an economy would not, by this criterion, count as much as a broader-based increase, even if the average increase were the same for both.

Todaro (1994) proposes a "distributive share index" or a "poverty-weighted index" as an alternative to the plain GDP or GNP, and the U.N. Human Development Report calculates "human development indicators." The inequality index I_F, such as calculated in (3) or (13), can also be used to measure the degree of economic development Y_D as a function of GDP, population, and inequality:

(23) $Y_D = Y_G / NI$

where Y_G is GDP. For a given per-capita GDP, the degree of economic development declines with increasing inequality.

If I_T is used,

(24) $Y_D = (Y_G / N)(1-G)$

As an example, in 1991 Mexico and Brazil had per-capita incomes of $2870 and $2930, with Gini coefficients .50 and .57, respectively (Schnitzer, 1994, pp. 259, 262). Hence, using I_T, Y_D for Mexico and Brazil are 1435 and 1260, Mexico having a higher development index despite its lower GDP due to the greater inequality in Brazil.

7. Conclusion

The equations I = CN and D = 1/C provide an integrated method of computing indexes for inequality, concentration, and diversification, given some distribution with N items. If C is an independently computed variable, then its use to compute inequality places a constraint on C, since certain intuitive constraints on I need to be met. Given N, I should increase with increasing concentration, and given C, I should increase with increasing N. For equal-share distributions, increasing N should decrease C proportionately so as to leave I equal. The Herfindahl-Hirschman and Tideman-Hall indexes satisfy these criteria. If I is to be calculated independently and C is the dependent variable, then the Tideman-Hall-Foldvary inequality index as a function of the Gini coefficient (14) has the desirable qualities needed for I = CN. As functions of Gini, the Tideman-Hall-Foldvary indexes for concentration (15), inequality (14) and diversification (16) may have many useful applications.

An inequality index I_F calculated from C_H or C_T in (1) or from G in (14) has some more consistent properties in relationship to N and C than straight Gini, so it merits empirical investigation as to its usefulness in computing inequality in income and wealth as a perhaps superior substitute for the raw Gini Coefficient. I don't disagree, however, with Hall and Tideman's (1967) statement that no best measurement may exist and that the measure should be suited to the use. The indexes presented here provide options worthy of investigation.

Endnote

1. See Maasoumi (1995) for a discussion of the Theil and Generalized Entropy indices and their applications.

References

Atkinson, Anthony. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2, no. 3 (September): 244-63.

Berrebi, Z. Moshe, and Silber, Jacques. 1985. "Income Inequality Indices and Deprivation: A Generalization." *Quarterly Journal of Economics* 100, no. 3 (August): 807-10.

Dagum, Camilio. 1987. "Gini ratio." In *The New Palgrave: A Dictionary of Economics*. Ed. John Eatwell, Murray Milgate, and Peter Newman. Vol. 2. London: MacMillan Press. Pp. 529-532.

Dalton, H. 1920. "The Measurement of Inequality of Incomes." *Economic Journal* 30 (September): 348-61.

Hall, Marshall, and Tideman, Nicolaus. 1967. "Measures of Concentration." *Journal of the American Statistical Association* 62 (March): 162-8.

Human Development Report. 1994. New York: Oxford University Press.

Maasoumi, Esfandiar. 1995. "Empirical Analyses of Inequality and Welfare." Forthcoming in *Handbook of Applied Microeconometrics*. Oxford: Basil Blackwell.

Sen, Amartya. 1973. *On Economic Inequality.* Oxford: Clarendon Press.

Shalit, Haim, and Yitzhaki, Shlomo. 1984. "Mean-Gini, Portfolio Theory, and the Pricing of Risky Assets." *Journal of Finance* 39, no. 5 (December): 1449-1468.

Schnitzer, Martin C. 1994. *Comparative Economic Systems.* Cincinnati: South-Western

Publishing.

Theil, Henry. 1967. *Economics and Information Theory.* Amsterdam: North-Holland.

Wright, Erik Olin. 1987. "Inequality." In *The New Palgrave: A Dictionary of Economics*. Eds. John Eatwell, Murray Milgate, and Peter Newman. Vol. 2. London: MacMillan Press. Pp. 815-19.