A Review of Concentration, Diversity or Entropy Metrics in Economics, Finance, Ecology and Communication Science

Isaac T. Tabner

April 2007

"There are fields of scientific work which have been explored from the different sides of pure mathematics, statistics, electrical engineering and neurophysiology in which every single notion receives a separate name from each group and which important work has been triplicated or quadruplicated, while still other important work is delayed by the unavailability in one field of results that have already become classical in the next field (Norbert Weiner, 1961, Cybernetics p. 2)".

Abstract

Many diversity, concentration and entropy metrics are analogous, sharing common origins in ecological science, engineering, mathematics, and communications theory. The difference in terminology depends upon the discipline in which the metrics are applied. The forthcoming article defines the concepts of concentration, diversity and entropy. The various indices used to measure these natural phenomena are examined and their origins identified. Selected examples of their application are reviewed. The aim is to provide a background to concentration metrics that have been applied in different fields of study, thus aiding the identification of topics where they might be used for future research.

Introduction

Diversity and concentration indices are analogous to one another, sharing common origins in ecological science, engineering, mathematics, entropy and communication theory. The only difference between the terms diversity and concentration is that the former is used in conjunction with evenness and equitability when the principles are applied in the ecological sciences, while the latter is used when referencing the same principles applied to research in economics. For example, in the context of an investment portfolio, concentration would be high if the majority of capital was invested in just a few relatively homogenous securities, while concentration would be low if the capital were distributed evenly between many heterogeneous securities. More specifically, stock market concentration refers to the degree to which a few disproportionately large firms dominate the returns of value weighted stock market indices such as the FTSE 100. The opposite situation would be low concentration, or a fragmented market, in which numerous small firms each accounted for a relatively small share of the overall market. In industrial economics, concentration is high when the majority of the output within an economy is accounted for by relatively few industries, or when the majority of output within an industry is accounted for by relatively few firms. In studies of ecology, diversity may refer to the degree to which the biomass in an ecosystem is distributed across many species of flora and fauna, as in a coral reef or tropical rainforest, or is concentrated in just a few species, as in a monoculture of Sitka spruce trees.

Examples of diversity metrics include, the Simpson Index, which can be traced to a paper published in Nature in 1949, by Simpson, titled "Measurement of Diversity". It is almost identical to the Hirschman-Herfindahl Index of industry concentration used in economics and developed independently by both Herfindahl (1950) and Hirschman (1964). The Simpson Index has also been referred to as the Yule Index after the similar measure Yule (1944) devised to characterise the vocabulary used by different authors. Yule (1944) is cited by Simpson (1949) as a key reference for his index, which is a combination of the ideas of Yule (1944) with those of Fisher et al (1943) and Williams (1946).

The entropy index of Shannon (1948), evolved from the the theory of communications engineering. Earlier researchers in this field include Nyquist (1924) and Hartley (1928). The objective of their research was to increase the efficiency of telegraphic information transmission. The Shannon Index is discussed further by Shannon & Weaver (1964) and later by Fernholz et al

(1998). In ecological texts the index is often referred to as the Shannon-Wiener Index in deference to Wiener (1961) who arrived at a similar index independently, in 1948. In fact, in the 1961 edition of Wiener's book "Cybernetics", Wiener cites the statistician Fisher, who is the same Fisher, cited as a key reference in an article by Simpson (1949), which details the Simpson index of diversity. Wiener also cites Shannon (1948) as a key developer of the index. Shannon and Weaver (1963) in turn refer to Wiener in their book that updates and develops some of the ideas outlined in Shannon (1948). Therefore, a relatively small group of statisticians, mathematicians and engineers, exerted a common influence in the development of the Simpson, Shannon and market concentration indices. Without reference to Shannon, Hart (1971) discusses entropy and other measures of concentration in the context of economics and business concentration. Diversity indices are also reviewed by Peet (1975) who cautions against the inappropriate scaling of diversity indices when conducting studies of community ecology.

Characteristics of concentration indices

At the generic level, absolute measures of concentration take into account both the number of different categories of units in a sample and the distribution of relative weights between these different categories. In an economic context, the units could be individual securities and the categories could define different firms or industries. In contrast, inequality measures of concentration only take into account the dispersion in the distribution of the weights between different categories and not the number of categories Clarke (1993). When comparing portfolios that have different numbers of constituents, or when the number of constituents in a portfolio or market is changing over time, there is an argument in favour of using absolute measures of concentration. An inequality measure of concentration allows comparison between samples of the same numerical size but when individual units account for different proportions of value or mass of the total sample.

Concentration Curve

The concentration curve is an absolute measure of concentration in which firm size inequality is represented by the convexity of the curve while firm numbers are indicated by the intersection of the curve at the 100% weight Clarke (1993), a hypothetical example is provided by Figure 1, which is recreated from Clarke (1993). The cumulative percentage weight is plotted on the y-axis against the number of firms starting from the largest on the x-axis. When concentration is at its lower limit, with weights equal for all firms, the line is straight. As concentration increases, the curve becomes more convex and moves further from the straight line, so that the shaded area in Figure 1 becomes larger. A curve that is very steep initially, and then flattens, represents a portfolio that is dominated by one or two large firms, but still contains many small firms that are, approximately, equal in size. On the other hand, if the concentration curve is smoother, large firms may still dominate, but the decrease in firm size, from largest to smallest is more continuous. Different concentration metrics place emphasis upon different parts of the concentration curve.

Lorenz curve

Figure 2 illustrates a hypothetical example of a Lorenz curve, also recreated from Clarke (1993). This differs from the concentration curve in that the cumulative percentage of the total number of firms are plotted from the smallest to the largest on the x-axis against the cumulative percentage of total market value on the y-axis. Because of the order reversal, a more concentrated market is indicated by greater concavity below the diagonal straight line rather than above it. Once again equal weighting, i.e. the lower limit of concentration is indicated by a straight line. As with the concentration curve, the Lorenz curve depicts the degree of concentration as the area between the curve and the intersecting straight line Clarke (1993).

Some absolute measures of concentration

The Simpson's index is calculated as in equation 1 and the value of D is an expression of the number of times one would have to take pairs of individuals at random from a population in order to select a pair of the same species Simpson (1949). In an economic context, D relates to the number

of times one would have to select pairs of firms at random from the population universe in order to find a pair in the same industry.

$$D = \frac{1}{\sum\limits_{i=1}^{N} w_i^2} \qquad E = \frac{1}{N} \times \frac{1}{\sum\limits_{i=1}^{N} w_i^2} \qquad \text{and}$$

(1)

E refers to the evenness of distribution in the population, sample, or portfolio. E can have a maximum value of one in a situation where species, or industries, are equally weighted. Values decreasing below one indicate increasing dominance by relatively few species, or industries. N is the number of species, or industries, in the population, while w is the population weight of each species, firm, industry, or other unit of measurement.

In cases where the sample or population size N is constant, the term $1/N$ can be deleted from the equation without loss of useful information. In some cases Simpson's index has been referred to as a dominance index as the squaring of the weights means that it is weighted towards the largest units in the sample.

Hirschman-Herfindahl Index

The Hirschman-Herfindahl Index, H, was initially attributed to Herfindahl (1950) in an unpublished dissertation for Columbia University. However, Hirschman (1964) later pointed out that an identical index, except for the presence of a square root, was in fact first published by himself in Hirschman (1945), although he noted that it has sometimes been incorrectly referred to as the Gini coefficient. The index usually referred to as the Herfindahl, or Hirschman-Herfindahl Index, is represented by equation 2 while the index claimed by Hirschman is represented by equation 3

$$H = \sum_{i=1}^{N} (x_i / x)^2 = \sum_{i=1}^{N} w_i^2$$

(2)

$$\sqrt{\sum_{i=1}^{N} w_i^2}$$

(3)

w_i is the weight of an individual company in a sample, x_i is the value of firm i and x is the total value of all firms i through N.

The H index takes account of all points on the concentration curve, as it is the sum of the squared weights of all firms in the industry. Therefore, if the total value of the market is defined as the sum of the market value of individual firms, then the share of the total market occupied by one firm is the market value of that firm divided by the value of the total market. This can be represented as follows: if n firms have market values xi ($i = 1,\ldots,n$) and the total value of the market is defined as x then

$$x = \sum_{i=1}^{n} x_i$$

and hence the share of the total market accounted for by the firm (w_i) is

$$w_i = \frac{x_i}{x}.$$

The squaring process results in the greatest weight being given to the larger firms in the industry or market constituent sample as in Simpson's Index D. It can be seen that H is simply the reciprocal of Simpson's index of Diversity D by examination of equation 4

$$H = \sum_{i=1}^{N}\left(x_i / x\right)^2 = \sum_{i=1}^{N} w_1^2 \quad \text{and} \quad D = \frac{1}{\sum_{i=1}^{N} w_i^2} \tag{4}$$

where the weights reflect the contribution made by each firm, industry, or species to the population universe.

Therefore,

$$\frac{1}{H} \equiv D \quad \text{while} \quad \frac{1}{D} \equiv H$$

with the result that for practical purposes the Hirschman-Herfindahl H index and the Simpson's E are merely derivations of one another. Publication of the Hirschman Index pre-dates the Simpson's paper by four years.

Clarke (1993) observes that the reciprocal of H is referred to as the numbers equivalent of H in that $1/H$ is equal to the number of firms in an equally weighted portfolio that would be required to produce the same value for H in the concentrated portfolio. Hence in the case of the FTSE 100 stock index, if H is equal to 0.05, then $1/0.05 = 20$ and thus twenty firms in an equally weighted portfolio would have the same value of H as is actually present in the 100 firm index portfolio. In other words, if the total number of firms is 100 and as few as 20 firms in the equally weighted portfolio will give the same value of H, then the portfolio is much more concentrated than if an equally weighted portfolio of 50 or even 100 firms were required.

If the average of firm size is taken together with the variance of firm size, H can be used to define a measure of inequality in the market values of different firms known as the coefficient of variation of firm size (c) represented by equation 5 (Clarke 1993)

$$c^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2 / \bar{x}^2 - 1 \quad \text{and} \quad H = \frac{c^2 + 1}{N} \tag{5}$$

Thus the H index depends upon the market share inequality as measured by c^2 and on the number of firms (N). H has an upper limit of unity in the hypothetical scenario in which the entire sample is represented by just one unit and the remaining units are infinitesimally small in size. It has a minimum value of $1/N \to 0$ in the case of many small equally sized firms. The square of c can effectively be regarded as a unit free measure of inequality (Clarke 1993).

Hannah and Kay's Index

The index of Hannah & Kay (1977), represented by equation 6, is very similar to the H index but it has the advantage of allowing the user to choose which part of the concentration curve to focus on by changing the value of α, an arbitrary elasticity parameter. This allows relatively greater weight to be given to large firms by increasing the value of α

$$R = \sum_{i=1}^{N} w_i^{\alpha} \quad \text{where } \alpha > 0 \tag{6}$$

The Hirschman-Herfindahl index is equivalent to R when $\alpha = 2$, while if α is sufficiently close to unity, the index will give the same ranking as that applied by the entropy index (E) discussed in the next section, Hannah & Kay (1977). The numbers equivalent to R is

$$R^{1/(1-\alpha)} \quad \text{or} \quad HR = \left(\sum_{i=1}^{N} w_i^{a} \right)^{1/(1-\alpha)}$$

and $\alpha > 0$ but $\alpha \neq 1$. This means that if, for example, there are 100 firms in a portfolio and $R = 5$, and $\alpha = 0.5$, then $1/(1-0.5) = 2$ and the numbers equivalent is equal to R^2. Hence the numbers equivalent is 25, meaning that in an equally weighted portfolio of just 25 firms would give the same

R of 5 that we found in the portfolio containing 100 firms. Note that if we set $\alpha < 1$, for example 0.5, then we are allowing the smallest firms to exert a relatively greater influence on the calculation of R than the largest firms.

Shannon's Diversity index, sometimes referred to as the Entropy Index

The Shannon Index (H_{sh}) is also often referred to as an entropy index (E). Shannon and Weaver (1963) provide a general definition of entropy when they cite the statement of Boltzmann (1894), as follows:

> *"Entropy is related to the missing information inasmuch as it is related to the number of alternative outcomes which remain possible to a physical system after all the macroscopically observable information concerning it has been recorded."*

In an ecological context the Shannon, or Entropy, Index simply questions how difficult it would be to predict correctly the species of the next individual sampled from a population of individuals. If uncertainty is low, the probability of a correct prediction increases. This implies that either the total number of species in the population is low or that the majority of individuals within the population are of the same species, i.e. diversity or evenness is low. On the other hand if the uncertainty number is high, the chance of a correct prediction is low implying that the total population is made up of a larger number of species and that they each represent a more evenly distributed share of the total biomass. In an economic context, this could represent the probability of a particular unit of output being derived from a particular firm, or a unit of capital in a portfolio being the security of a particular firm, when the units of the output or units of portfolio capital are sampled from the whole economy and the market portfolio respectively. The Index is calculated using equation 7

$$E = -\sum_{i=1}^{N} w_i \ln w_i \quad = \quad E = \sum_{i=1}^{N} w_i \ln(1/w_i)$$

(7).

The entropy index (E) is also related to Shannon's equitability index J as depicted in

$$J = \frac{E}{E_{max}} = \frac{-\sum_{i=1}^{N} w_i \ln w_i}{\ln N}$$

(8)

In an ecological context where E and E_{max} are diversity indices, E_{max} is the maximum diversity that could be found in a situation where all species were equally abundant. It will give a value identical to Hannah and Kay's R when α is set to equal unity (Clarke 1993). The minus sign in the numerator of the expression, reflects the fact that any probability is a number less than or equal to unity, and the logarithms of numbers less than one are negative. Thus a minus sign is used here in order to ensure that Shannon's E is positive.

The Divergence Index of Kacperczyk, Sialm and Zheng

Kacperczyk et al (2005) examined the correlation between actively managed mutual fund portfolio concentration, measured using their divergence index (DI), and abnormal returns. Their index of divergence is related to the Hirschman-Herfindahl H index and is calculated as

$$DI_t^F = \sum_{i=1}^{N} \left(w_{i,t}^F - w_{i,t}^M\right)^2$$

,

(9)

where $w_{i,t}^F$ is the weight of the asset i in the *fund* portfolio and $w_{i,t}^M$ is the weight of the asset i in the *market* portfolio.

The basic function of the divergence index is to measure the degree to which the capital allocation of an individual actively managed fund portfolio differs from the capital allocation that would result from a passive market-tracking portfolio.

Inequality measures of concentration

Inequality measures of concentration ignore the absolute numbers of firms in a sample and can thus be viewed as summary representations of the Lorenz curve in the same way that absolute measures summarise the concentration curve (Clarke 1993).

Concentration ratio

Clarke (1993) defines the concentration ratio, *C,* as the proportion of industry output accounted for by the *r* largest firms or, the combined weight of the *r* largest firms in an index, when *r* is an arbitrary number. The concentration ratio *C* is easy to calculate as in

$$C_r = \sum_{i=1}^{r} \frac{x_i}{x} = \sum_{i=1}^{r} w_i$$

(10)

and easy to understand. However, it only focuses on a single part of the concentration curve and if the shape of the concentration curve changes over time, comparison of the C_r values in a time series will provide an incomplete picture of the distribution of firm size in an index.

Coefficient of variation

The coefficient of variation has already been discussed in relation to the Hirschman-Herfindahl H index above. It is the ratio of the standard deviation of firm size to the mean of firm size. Hence it is a unit free measure of dispersion and hence firm size inequality represented by equation 5.

Variance of the logarithms of firm size

The variance of the logarithms of firm size, V^2, is particularly useful if the distribution of firm size in an index approximates to the log normal distribution, something that frequently occurs in practice. The index is calculated as

$$V^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\log\left(x_i / \bar{x}_g \right) \right]^2$$

(11)

where $\bar{x}_g$ is the geometric mean of firm size. In the situation where firm values are approximately log normal and Lorenz curves are non-intersecting, V^2 provides an unambiguous ranking of firm size inequality Clarke (1993).

The Gini coefficient

The Gini coefficient, *G*, first proposed by Gini (1912), *G* is a ratio of the shaded area between the Lorenz curve and the diagonal straight line, in figure 2, and the area of the triangle falling below the diagonal straight line. The Gini coefficient is calculated as the relative mean difference i.e., the mean of the difference between every possible pair of individuals, x_i, x_j, in a matrix divided by the mean of the firm size, as in

$$G = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} |x_i - x_j|}{2N^2 \mu}$$

(12)

where *N* is the number of firms and μ is the mean of the firm size.

The Gini coefficient has a minimum value of zero, when all firms are equal in size and a theoretical maximum value of unity, when every firm except one has a weight of zero. Gini's original paper is in Italian, but an English explanation is provided by the online encyclopaedia of mathematics, Mathworld.com edited by Weisstein (1999-2004).

Selected studies in economics and finance that apply concentration metrics

A number of studies have examined industrial concentration at both the macro and sector level of the economy. However, there are relatively few studies that have measured concentration using the equity market value, i.e. market capitalisation, of firms listed on stock exchanges. One such example, Roll (1992) examines concentration in equity market capitalisation accross different country's stock market indices and concludes that more concentrated stock indices tend to exhibit greater volatility of returns. Tabner (2006) examines changes in concentration of the FTSE 100 Index of leading UK firms over the period 1984 – 2004 and concludes that there is little evidence that increases in UK market concentration has led to increases in the volatility of UK equity market returns.

Hou and Robinson (2006) use the Hirschman Herfindahl index to measure concentration of sales in industry product markets when they examine the relationship between industry concentration and the returns of industry constituent firms in the US market over the period from 1963 to 2001. They find that firms in more concentrated industries exhibit lower returns and have higher dividend payout ratios than firms in less concentrated and potentially more competitive industries after adjusting for standard risk factors document in other asset pricing studies. They argue that firms operating in less concentrated industries are likely to be participating in and exposed to more rapid product innovation making their cashflows riskier than in more concentrated industries where barriers to entry are likely to be higher and innovation lower. Hence the return premium arising from less concentrated more competitive industries is a compensation for the increased risk bourne by investors in those industries.

Kacperczyk et al (2005) examine the concentration of actively managed US equity portfolios with the relative investment performance of those funds over the period December 1984 to December 1999. The authors use their divergence index, *DI*, to measure the degree to which the asset allocation of an actively managed mutual fund portfolio diverges from the asset allocation of the market portfolio. They suggest that active managers who diverge the furthest from the market portfolio generally, but not always, maintain portfolios that are more concentrated than the market portfolio. In addition, they propose that the greater the degree of divergence from the market portfolio, the more aggressive, and the more confident, is the active management. They argue that portfolio managers which adopt the strategy of diverging from the asset allocation of the market portfolio do so, because they have a high degree of self confidence regarding their stock picking, asset allocation ability and information advantage in relation to the market as a whole. They find that the added confidence implicit in the managers of more concentrated funds appears to be justified as concentrated funds produce better risk and style adjusted returns, on average.

Hannah & Kay (1977) in their book, "Concentration in Modern Industry" (pp. 18 – 19) review regression studies that relate concentration to profitability. Notable among these are Bain (1951) who studied a number of industries and observed that industries in which the largest eight firms held combined market shares in excess of 70% earned markedly higher returns. Kilpatrick (1967) finds that levels of profits and changes in profits relate to both the concentration ratio and a dummy variable reflecting whether or not a critical level of concentration is exceeded. Meehan and Duchesnau (1973) also found results consistent with those of Bain (1951). Hart & Clarke (1980) present results from an extensive study that they undertook of the UK market.

In the light of recent developments, it is interesting to note that Hannah & Kay (1977) in their study of concentration in British industry provided the following insight thirty years ago

"If there is another merger wave comparable to that of the 1920s or 1960s, then apparently fanciful predictions such as those of Newbould & Jackson (1972) of a hundred firms controlling 70 or 80 percent of all manufacturing activity could be fulfilled within ten or fifteen years (Hannah and Kay 1977 p. 114)."

Hannah and Kay precede this statement with a suggestion that merger activity is a primary cause of increasing concentration and argue that small firms will only achieve sufficient growth to counter the trend towards increased concentration, if, for some reason future merger activity is very low. They follow on with the comment that at the time of writing, just prior to 1977, current merger activity was indeed at historically low levels and if it did not recover then the concentration level, predicted in the above quote, would not be reached until the next century. By the year 2000,

only ten firms accounted for 42% of the total £1.8 trillion market capitalisation of the UK stock market and the largest company, Vodafone, represented 9% of the total UK market capitalisation. This followed a series of mergers including: the formation of Glaxo-Welcome, which later became Glaxo-Smithkline, BP with Amoco and the combination of Vodafone with Mannessmann. Clearly the "apparently fanciful prediction" occurred within 25 or so years, although not within the ten or fifteen suggested by Newbould and Jackson (1972), largely due to renewed phases of intense merger activity, as predicted by Hannah and Kay.
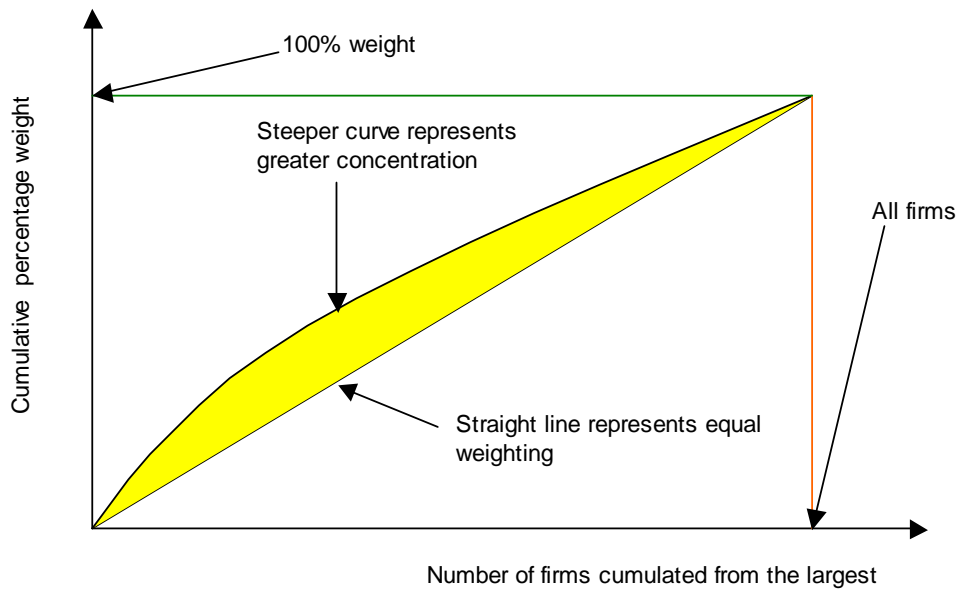
Conclusion

This paper has examined a range of concentration and diversity metrics that have been applied by researchers in economics, finance, ecology and communications engineering. The origin and history of the metrics is documented. The strengths, weakness and origins of the concentration indices have also been identified. Selected studies that applied concentration metrics in finance and ecomomics are examined. By providing an entry to prior research, the paper will benefit researchers from various disciplines who wish to investigate the implications of concentration, diversity and eveness of population and sample distributions.

REFERENCES

Bain, J.S. (1951). Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-40. Quarterly Journal of Economics 65:3 pp 293-324.

Clarke R. (1993). Industrial Economics, Cambridge Massachusetts: Blackwell.

Fernholz, R., Garvy, R. and Hannon, J. (1998). Diversity-Weighted Indexing. A New Approach to Passive Investing. The Journal of Portfolio Management 24:2 pp 74-82.

Fisher, R.A., Corbet, S.A. and Williams, C.B. (1943). The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population . Journal of Animal Ecology 12:1 pp 42-58.

Gibrat R. (1931). Les Inegalites Economiques, Partly Translated and Reprinted As "On Economic Inequalities", Paris: Reprint "International Economic Papers (1957)".

Gini, C. (1912). Variabilita' e Mutabilata. Reprinted in *Memorie di metodologia statistica* (Ed. AV. Pizetti and T. Salvemini.) Rome: Libreria Eredi Virgilio Veschi, 1955. Cited by Weisstein, E. W. Mathworld.com (1999-2004)

Hannah L. & Kay, J. A. (1977). Concentration in Modern Industry. Theory, Measurement and the U.K. Experience, London and Basingstoke: The Macmillan Press Ltd.

Hart, P.E. (1971). Entropy and Other Measures of Concentration. Journal of the Royal Statistical Society. Series A (General) 134:1 pp 73-85.

Hart P.E. & Clarke, R. (1980). Concentration in British Industry 1935-75, Cambridge: Cambridge University Press.

Hartley, R.V.L. (1928). Transmission of Information. Bell System Technical Journal 7:4 pp 535-563.

Herfindahl, O.C. (1950). Concentration in the US Steel Industry. Doctoral Dissertation Unpublished, Columbia University

Hirschman A.O. (1945). National Power and the Structure of Foreign Trade, Berkeley:

Hirschman, A.O. (1964). The Paternity of an Index. The American Economic Review 54: pp 761-2.

Hou, K. and Robinson D. T. (2006). Industry Concentration and Average Stock Returns. Journal of Finance. 61:4 (August) pp 1927 -1956.

Kacperczyk, M., Sialm, C. and Zheng, L. (2005). On the Industry Concentration of Actively Managed Equity Mutual Funds. Journal of Finance, Vol. 60, No. 4, pp 1983 – 2011.
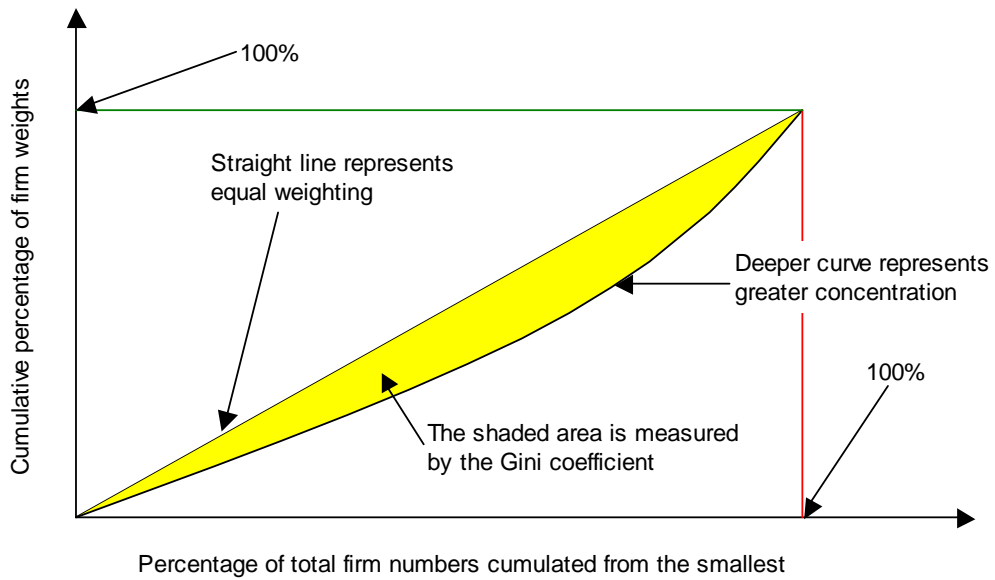
Kilpatrick, R.W. (1967). The Choice Among Alternative Measures of Industrial Concentration. Review of Economics and Statistics 49:2 pp 258-260.

Meehan, J.W. & Duchesnau, T.D. (1973). The Critical Level of Concentration. Journal of Industrial Economics 22:1 pp 21-36.

Newbould G.D. & Jackson, A. S. (1972). The Receding Ideal, Liverpool: Guthstead Ltd .

Nyquist, H. (1924). Certain Factors Affecting Telegraph Speed. Bell System Technical Journal 3:4 pp 324-346.

Peet, K. (1975). Relative Diversity Indices. Ecology 56:2 (Spring) pp 496 – 498.

Roll, R. (1992). Industrial Structure and the Comparative Behaviour of International Stock Market Indices. Journal of Finance 47:1 pp 3-41.

Shannon, C.E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal 27:3 pp 379-423.

Shannon C.E. & Weaver, W. (1964). The Mathematical Theory of Communication, Urbana & Chicago: University of Illinois Press.

Simpson, E.H. (1949). Measurement of Diversity. Nature 163:13 pp 688.

Tabner, I.T. (2006). Benchmark Concentration: Value Weights versus Equal Weights in the FTSE 100 Index. SSRN Working Paper.

Weisstein, E.W. (1999-2004). Gini Coefficient. Mathworld

Wiener N. (1961). Cybernetics, or Control and Communication in the Animal and the Machine., New York & London: M.I.T. Press & John Wiley & Sons, Inc.

Williams, C.B. (1946). Yule's Characteristic and the Index of Diversity. Nature 157 :13 pp 482.

Yule G.U. (1944). Statistical Study of Literary Vocabulary, Cambridge: Cambridge University Press.

Figure 1   Concentration curve

From Clarke (1993)

Figure 2   Lorenz curve of a hypothetical portfolio

From Clarke (1993)