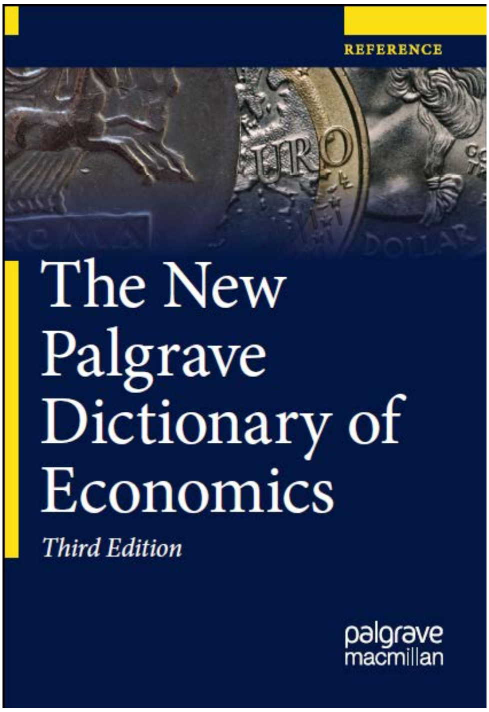
Gini Ratio

Camilo Dagum



In a national economy, the price system determines both resource allocation and the income distribution. The imputation to the factors of production of the mass of income associated with an economy's output determines its distribution by factor shares, or *functional* income distribution. This mainstream of research follows Ricardo's (1817) contribution. Another mainstream of research was initiated by Pareto (1895, 1897), and deals with the distribution of a mass of income among the members of a set of economic units (family, household, individual), considering either the total income of each economic unit or its disaggregation by source of income, such as wages and salaries, property income, self-employment income, transfers, etc. This type of inquiry deals with distribution by size of income, or *personal* income distribution, and the quantitative assessment of the relative degree of income inequality among the members of a given set of economic units. Such inquiries provide basic quantitative information in support of a comprehensive research strategy on income distributions, including causal explanations for social welfare and policy.

It is of interest to remark that Pareto's research on income distribution was motivated by the polemic he engaged in with French and Italian socialists concerning the ways and means of achieving a less unequal distribution. Thus, the actual *measurement* of inequality was brought to the fore, with its main purposes the assessment of (i) the evolution of inequality in a given country or region, and (ii) the relative degree of inequality between countries or regions.

In a series of methodological and applied contributions Corrado Gini (1955) enriched this field of research. In 1910 he corrected the interpretation of Pareto's inequality parameter and, in 1912, proposed a new measure of income inequality, the Gini ratio.

Pareto (1896, 1897) specified three versions of this model of income distribution. The most widely used model is Pareto Type I

$$S(x) = 1 - F(x) = (x/x_0)^{-\alpha}, \quad 0 < x_0 < x, \alpha > 1, \tag{1}$$

where $S(x) = P(X > x)$ is the survival distribution function (SDF) of the income variable X, $F(x)$ is the cumulative distribution function (CDF), x_0 is the minimum value of X, α is a scale-free inequality parameter, and the mathematical expectation of income is

$$\mu = E(X) = \alpha x_0^\alpha \int_{x_0}^{\infty} x^{-\alpha} dx = \alpha x_0/(\alpha - 1). \tag{2}$$

Pareto seems to have assumed that income growth implies less income inequality. This assumption, together with eqn (2), led him to the conclusion that income inequality is an increasing function of α. Gini (1910) reversed this interpretation, proving that, given model (1), income inequality is a decreasing function of α. Gini's rationale was as follows: given n units with incomes $x_1 \leq x_2 \leq x_3, \ldots, \leq x_n$, the average of the last $m(m \leq n)$ income units $\Sigma_{i=0}^{m-1} x_{n-i}/m$ is greater than or equal to the average income $\mu = \Sigma_{i=0}^{n} x_i/n$ of the population, hence, there exists a $\delta \geq 1$ such that

$$\left.\sum_{i=0}^{m-1} x_{n-i} \Big/ \sum_{i=1}^{n} x_i\right)^{\delta} = m/n, \quad \delta \geq 1, \tag{3}$$

Equation (3) is known as the Gini model. Gini (1910) interpreted the scale-free parameter δ as a measure of income inequality and called it a *concentration ratio* because it is an increasing function of the concentration of income in the upper income groups. For this reason, Gini called eqn (3) a *concentration curve,* where the abscissa represents the CDF $F(x_m) = m/n$ and the ordinate the income share $\sum_{i=1}^{m} x_i / \sum_{i=1}^{n} x_i$, $m = 1, 2, \ldots, n$, δ being an unknown parameter that has to be estimated.

Using the CDF $F(x)$ and the Lorenz curve $L(x)$ (also called the Lorenz-Gini curve since it was independently introduced by both authors), eqn (3) takes the form

$$1 - F(x) = [1 - L(x)]^{\delta}, \quad \delta \geq 1, \tag{4}$$

where

$$L(y) = (1/\mu) \int_x^y x \mathrm{d}F(x). \tag{5}$$

Replacing $F(x)$ from model (1) into eqns (4) and (5), Gini (1910) proved that $\delta = \alpha/(\alpha - 1)$ and thus reversed Pareto's interpretation of α. In fact, when $\alpha \to \infty$, $\delta \to 1$ and $F(x) = L(x)$, and the mass of income is equally distributed.

Gini (1912) specified the Gini mean difference with and without replacement. The latter is by definition

$$\Delta = \sum_{j=1}^{n} \sum_{i=1}^{n} |x_j - x_i|/n(n-1),$$
$$0 \leq \Delta \leq 2\mu, \tag{6}$$

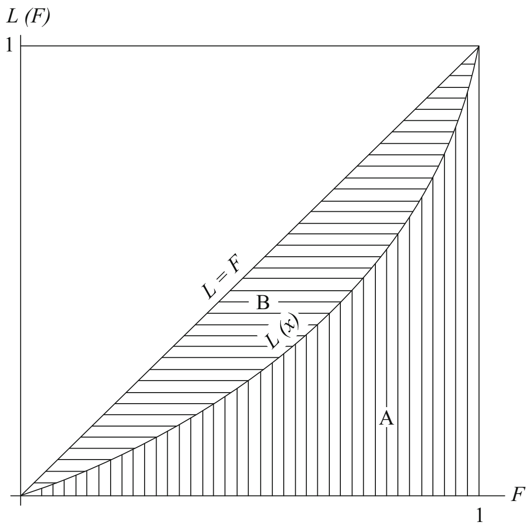
and using the Riemann-Stieltjes integral, which covers, as particular cases, both discrete and continuous distributions, we have

$$\Delta = \int_0^{\infty} \int_0^{\infty} |y - x| \mathrm{d}F(x) \mathrm{d}F(y), \tag{7}$$

where X and Y are identically and independently distributed variables. When $x_1 = x_2 = \ldots = x_n$, $\Delta = 0$, and when $x_1 = x_2 = \ldots = x_{x-1} = 0$ and $x_n = n\mu$ (the total income), $\Delta = 2\mu$.

Since Δ is a monotonie increasing function of the degree of income inequality, Gini (1912) specified

$$G = \Delta/2\mu, \quad 0 \leq G \leq 1 \tag{8}$$



Gini Ratio, Fig. 1 Lorenz curve $L(x)$ and Gini ratio G

as an income inequality measure. Equation (8) is known as the Gini ratio or Gini index and it is widely used in theoretical and applied research on income and wealth distributions.

Gini (1914) proved the important theorem that $G = \Delta/2\mu$ is equal to twice the area between the equidistribution line $F(x) = L(x)$ and the Lorenz curve $L(x)$ (see Fig. 1). Moreover,

$$\begin{aligned} G = \Delta/2\mu &= 2\int_0^1 (F-L)\mathrm{d}F \\ &= (2/\mu)\int_0^\infty x\left[F(x)-\frac{1}{2}\right]\mathrm{d}F(x) \\ &= (2/\mu)\int_0^\infty x\left[\frac{1}{2}-S(x)\right]\mathrm{d}F(x). \end{aligned} \tag{9}$$

For the discrete case, it follows from eqns (6) and (8), that

$$\begin{aligned} (F(x)) &= (1/\mu)\int_0^x y dF(y) G = 2B = 1-2A = 1-1\int_0^1 L dF \\ G &= [2/n(n-1)\mu]\sum_{k=1}^n k x_k - (n+1)/(n-1) \\ &= (n+1)(n-1) - [2/n(n-1)\mu]\sum_{k=1}^n (n-k+1)x_k, \end{aligned} \tag{10}$$

showing that the welfare function underlying the Gini ratio is a rank-order-weighted sum of the economic units' income shares.

The properties that an income inequality measure must fulfil were first discussed by Dalton (1920). It can be shown (Dagum 1983, pp. 34–5) that G fulfils the properties of (i) transfer, (ii) proportional addition to incomes, (iii) equal addition to incomes, (iv) proportional addition to persons, (v) symmetry, (vi) normalization, and (vii) operationality.

The Gini ratio is sensitive to transfers to all income levels. In fact, it follows from eqn (10), that a transfer of h dollars from the richer j to the poorer i theorem, without modifying their income ranks, is

$$\Delta G(j,i;h) = -2(j-i)h/n(n-1)\mu > 0, \qquad j > i, \tag{11}$$

therefore $-\Delta G$ is an increasing function of $j - i = F(x_j) - F(x_i)$ and a decreasing function of both n and μ. The maximum reduction of G is achieved when $h = (x_j - x_i)/2$, and is not necessarily given by eqn (11) unless the transfer fulfils certain conditions with respect to the original income ranking of the population.

Often, the Gini ratio is misinterpreted when it is incorrectly claimed that it attaches more weight to transfers to income near the mode of the distribution than at the tails. In particular, the misinterpretation arises when eqn (11) instead of eqn (9) is applied to unimodal distributions when assessing the relative sensitivity of G to income transfers. Consequently, the assumptions supporting the mathematical structure of eqn (11) are ignored.

It follows from eqn (10) that the Gini ratio fulfils the duality principle between the representation of an inequality measure (I) satisfying the principle of transfer, i.e. $I = E[V(x)]$, and that of a social welfare (SW) function, i.e. $SW = E[-V(x)]$, where $-V(x)$ is concave, or more generally, S-concave (Berge 1966). It follows from eqn (9), that two equivalent forms of $V(x)$ in $G = E[V(x)]$ are

$$V(x) = 2xF(x)/\mu - 1, \quad \text{and} \quad V(x) = x[2F(x) - 1]/\mu. \tag{12}$$

Sen (1974) introduced an axiomatic system for the SW interpretation of the Gini ratio based on the individual income ranking of the population suggested by the structure of eqn (10). Following Sen's ideas, Kakwani (1980, pp. 77–9) presented a SW interpretation of the Gini ratio as a function of income. Both approaches can be presented in a compact form by making use of the SDF $S(x) = 1 - F(x)$ and the first moment survival distribution function $S_1(x) = 1 - L(x)$. In fact, specifying the SW function

$$\mathrm{SW}(X) = E[Xv(x)] \tag{13}$$

where $v(X)$ is a decreasing and differentiable function of X, and making $v(X) = 2S(X) = 2(1 - F(X))$, i.e., twice the frequency of economic units with income greater than X, we deduce

$$\mathrm{SW}(X) = 2\int_0^\infty xS(x)\ \mathrm{d}F(x) = \mu(1 - G), \tag{14}$$

which proves Sen's (1974, p. 410) theorem that the SW function (14) ranks a set of distributions of a constant total income and population in precisely the same way as the negative of the Gini ratio of the respective distributions, i.e. in reverse order from that by the cardinal value of the Gini ratio. On the other hand, making $v(X) = bS_1(X) = b[1 - L(X)]$, $b > 0$ and $\int_0^\infty v(x)\ \mathrm{d}F(x) = 1$, where $S_1(x)$ is the income share of the economic units with income greater than x, we deduce

$$b\int_0^\infty [1 - L(x)]\mathrm{d}F(x) = b(1 + G)/2 = 1, \text{and}$$
$$\mathrm{SW}(X) = [2/(1 + G)]\int_0^\infty x[1 - L(x)]\mathrm{d}F(x) = \mu/(1 + G), \tag{15}$$

which also states that the SW function (15) is a decreasing function of the Gini ratio. The result obtained in eqn (14) supports Sen's (1976, p. 384) cogent statement that 'one might wonder about the significance of the debate on the non-existence of any additive utility function which ranks income distributions in the same order as the Gini ratio'.

The Gini ratio stimulated important contributions such as:

(i) The construction of a confidence interval for G. Given a random sample of size n, eqn (10) is an unbiased estimator of G. However, income distribution data are presented by class intervals, hence Gini (1914) proposed the formula $G_L = 1–2A$, where A is the area under the Lorenz curve (Fig. 1) estimated by application of the trapezoidal approximation to

$$\int_0^1 L\ dF,$$

thus underestimating G because the trapezoidal rule implies that within each interval, income is equally distributed. Gastwirth (1972) derived an upper bound G_u, by maximizing the spread within each income interval, and proposed (G_L, G_u) as a confidence interval within which a parametric estimate of G should fall. Dagum (1980a) proved that his confidence interval is a necessary but not sufficient condition to assess a model goodness of fit.

(ii) The Gini ratio gives a welfare ranking (weak ordering) of a set of income distributions of a constant mass of income and over a constant population, and a strict partial ordering among the subset of income distributions with non-intersecting Lorenz curves. This conclusion is further supported by eqns (14) and (15).

(iii) The welfare ranking of income distributions with equal and different means can be obtained via a decision function $R(G, D)$, where the ratio G states the preference for less inequality (inequality aversion) regardless of the mean income (so that the partial derivative $R_G < 0$), and the relative economic affluence D (Dagum, 1980b, 1987) states the preference for more income (poverty aversion), so that the partial derivative $R_D > 0$.

(iv) Research on the economics of poverty led Sen (1976) to the specification of an axiomatic structure of a new poverty measure as a function of (a) the relative frequency of the poor members of the population, (b) a weighted average of the poverty gap, i.e. the aggregate shortfall from the poverty line of the poor population, and (c) the Gini ratio of the income distribution of the subpopulation with incomes below the poverty line.

(v) Gini (1932) introduced a new coordinate system taking as the abscissa the egalitarian line $F = L$ and as the ordinate the distance between the Lorenz curve and the egalitarian line. Gini thoroughly analysed this new coordinate system and its relation to the G ratio. Kakwani (1980, ch. 7) worked with a similar transformation.

(vi) Analysing consumer behaviour in India, Mahalanobis (1960) extended and generalized the Lorenz curve and the Gini ratio with the introduction of the concentration curve and ratio, respectively. Other authors such as Kakwani (1980, chs 8–14) made further contributions and dealt with the relationships among the distribution of several economic variables such as expenditures and income after tax, and investigated the degree of tax and public expenditure progressivity or regressivity. If $y = g(x)$ is the function of income that is the object of inquiry, $g(x)$ must be non-negative. For the particular case of $g(x) = x$, the concentration curve and ratio are identical to the Lorenz curve and Gini ratio, respectively. Moreover, if $g(x)$ is an increasing and differentiable function of x, i.e. $g'(x) > 0$, then the concentration ratio is equal to the Gini ratio for the function $g(x)$.

(vii) The decomposition approach disaggregates a population according to some relevant socio-economic attributes and analyses the equality within each subpopulation and between them, and assesses the contribution of each subpopulation to overall inequality. This approach also disaggregates the income variable by source of income such as wages and salaries, self-employment, pension and government transfers. Bhattacharya and Mahalanobis (1967) were the first to deal with the decomposition of the Gini ratio. Several authors made further contributions to this topic, among them Pyatt (1976) and Shorrocks (1983).

Bibliography

Berge, C. 1966. *Espaces topologiques, fonctions multivoques*, 2nd edn. Paris: Dunod.

Bhattacharya, N., and P.C. Mahalanobis. 1967. Regional dispartities in household consumption in India. *Journal of the American Statistical Association* 62: 143–161.

Dagum, C. 1980a. The generation and distribution of income, the Lorenz curve and the Gini ratio. *Economie Appliquée* 33 (2): 327–367.

Dagum, C. 1980b. Inequality measures between income distributions with applications. *Econometrica* 48 (7): 1791–1803.

Dagum, C. 1983. Income inequality measures. In *Encyclopedia of statistical sciences*, ed. S. Kotz and N.L. Johnson, vol. IV, 34–40. New York: Wiley.

Dagum, C. 1987. Measuring the economic affluence between populations of income receivers. *Journal of Business and Economic Statistics* 5 (1): 5–12.

Dalton, H. 1920. The measurement of the inequality of incomes. *Economic Journal* 30 (September): 348–361.

Gastwirth, J.L. 1972. The estimation of the Lorenz curve and Gini index. *Review of Economics and Statistics* 54: 306–316.

Gini, C. 1910. Indici di concentrazione e di dependenza. *Atti delta III Riunione delta Società Italiana per il Progresso delle Scienze*, in Gini (1955), 3–120.

Gini, C. 1912. Variabilitàe mutabilità. *Studi economico-giuridici, Università di Cagliari* III, 2a, in Gini (1955), 211–382.

Gini, C. 1914. Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del R. Istituto Veneto di Scienze, Lettere ed Arti,* in Gini (1955), 411–459.

Gini, C. 1932. Intorno alle curve di concentrazione. *Metron* 9(3–4), in Gini (1955), 651–724.

Gini, C. 1955. *Memorie di metodologia statistica.* Vol. 1: *Variabilità > concentrazione,* ed. E. Pizetti and T. Salvemini. Rome: Libreria Eredi Virgilio Veschi.

Kakwani, N. 1980. *Inequality and poverty: Methods of estimation and policy applications*. Oxford: Oxford University Press.

Mahalanobis, P.C. 1960. A method of fractile graphical analysis. *Econometrica* 28 (2): 325–351.
Pareto, V. 1895. La legge della domanda. *Giornale degli Economisti* 5: 59–68.
Pareto, V. 1896. Ecrits sur la courbe de la répartition de la richesse. In *Oeuvres complètes de Vilfredo Pareto*, ed. Giovanni Busino. Geneva: Librairie Droz, 1965.
Pareto, V. 1897. *Cours d' éeconomie politique,* ed. G.H. Bousquet and G. Busino. Geneva: Librairie Droz, 1964.
Pyatt, G. 1976. On the interpretation and disaggregation of Gini coefficients. *Economic Journal* 86 (June): 243–255.
Ricardo, D. 1817. Principles of political economy. In *Works and correspondence of David Ricardo*, ed. P. Sraffa, vol. I. Cambridge: Cambridge University Press. 1951.
Sen, A.K. 1974. Information bases of alternative welfare approaches. *Journal of Public Economics* 3: 387–403.
Sen, A.K. 1976. Poverty: An ordinal approach to measurement. *Econometrica* 44 (2): 219–231.
Shorrocks, A.F. 1983. The impact of income components on the distribution of family income. *Quarterly Journal of Economics* 98 (2): 311–326.

Gini, Corrado (1884–1965)

Camilo Dagum

Gini, perhaps best known to economists because of the Gini Coefficient, was born in Motta di Livenza, Italy and died in Rome. He studied at the University of Bologna; his doctoral thesis *Il sesso dal punto di vista statistico* (1908), defended in 1905, was awarded the Vittorio Emanuele prize for social sciences. Gini distinguished himself as a teacher and a researcher. In 1909 he was appointed an assistant professor of the University of Cagliari, becoming full professor a year later. Gini won a chair at the University of Padova in 1913, then joined the University of Rome in 1925, where in 1955 he was awarded the distinction of emeritus professor. Social scientist and statistician, Gini taught economics, statistics, sociology and demography, making path-breaking contributions to these highly related disciplines. Among them we mention the neo-organicist theory (Gini 1909, 1924a) that presents a dynamic theory of society in which demographic factors (differential birth rates among social classes and social mobility) play a basic role. In this theory, Gini introduced and analysed self-conservation, self-regulative and self-re-equilibrating mechanisms, thus offering a well-structured anticipation of Wiener's cybernetics, von Bertalanffy's general system theory and modern disequilibrium economics. He provided new insights to the analysis of inter- and intra-national migrations (Gini 1948) and demographic dynamics (Gini 1908, 1909, 1912a, 1931). He developed a methodology to evaluate the income and wealth of nations (Gini 1914a, 1959) including a discussion of human capital, already present in his research on the causes and consequences of international migrations. In this context he specified a model of income and wealth distributions and a measure of income and wealth inequalities (Gini 1909, 1912b, 1914b, 1955). Gini's research interests motivated important contributions to statistics and economics, such as the Gini identity (1921, 1924b) on price index numbers, the Gini mean difference (1912b), the transvariation theory (Gini 1916, 1960), the index of dissimilarity (Gini 1914c) and the Gini Coefficient. Gini founded several scientific journals, such as *Metron* and *Genus*, and academic institutions, such as the Institute and Faculty of Statistics, Demography and Actuarial Sciences of the University of Rome; and was the organizer and first president (1926–1932) of the Istituto Centrale di Statistica. An extraordinarily prolific writer and thinker, endowed with powerful new ideas that he developed in more

G

than 70 books and 700 articles, Gini was in the 20th century a true Renaissance man.

Selected Works

1908. *Il sesso dal punto di vista statistico*. Milan: Sandrom.

1909. Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza. *Giornale degli Economisti* 38: 27–83.

1912a. *I fattori demografici dell'evoluzione delle nazioni*. Turin: Bocca.

1912b. Variabilità e mutabilità. Reprinted in Gini (1955).

1914a. *L'ammontare e la composizione della ricchezza delle nazioni*. 2nd ed., Turin: UTET, 1962.

1914b. Sulla misura della concentrazione e della variabilità dei caratteri. Reprinted in Gini (1955).

1914c. Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche. *Atti del R. Istituto Veneto di Scienze, Lettere ed Arti* 74, Part II.

1916. Il concetto di transvarizione e le sue prime applicazioni. Reprinted in Gini (1960).

1921. Sull'interpolazione di una retta quando i valori della variabile indipendente sono affetti da errori accidentali. *Metron* 1:63–82.

1924a. *Patologia economica*, 5th ed. Turin: UTET, 1954.

1924b. Quelques considérations au sujet de la construction des nombres indices des prix et des questions analogues. *Metron* 4:3–162.

1931. *Le basi scientifiche della politica della popolazione*. Catania: Studio Editoriale Moderno.

1948. Apparent and real causes of American prosperity. *Banca Nazionale del Lavoro Quarterly Review* 6, July–September.

1955. *Memorie di metodologia statistica*. Vol. 1: *Variabilità e concentrazione*, ed. A. Pizzetti and T. Salvemini. Rome: Veschi.

1959. *Ricchezza e reddito*. Turin: UTET.

1960. *Transvariazione*, ed. G. Ottaviani. Rome: Libreria Goliardica.