ELSEVIER

Economics Letters 56 (1997) 45–49

economics letters

A simple way to calculate the Gini coefficient, and some implications

Branko Milanovic*

World Bank, Washington, D.C., USA

Received 7 October 1996; accepted 19 March 1997

Abstract

The paper proposes a new, and a much simpler, way to calculate the Gini coefficient. The existence of a relationship between linear (concave or convex) Pen's income parade, and specific values of the Gini coefficient is derived. © 1997 Elsevier Science S.A.

Keywords: Personal income distribution; Gini coefficient; Pen's parade

JEL classification: D31

There are many ways to express and calculate the Gini coefficient from the individual data.[1] A popular and easy formula (derived in Pyatt et al. (1980)), and used more recently by Lerman and Yitzhaki (1984) and Yitzhaki (1994)) is:

$$G = \frac{2\,\text{covar}(y, r_y)}{N\bar{y}}, \tag{1}$$

where covar(y, r_y) is the covariance between income (y) and ranks of all individuals[2] according to their income (r_y) ranging from the poorest individual (rank = 1) to the richest (rank = N). N is total number of individuals, and $\bar{y}$ = mean income.

The purpose of this note is to propose an even simpler way to calculate the Gini coefficient and to explore some implications of this approach.

We start from expression (1). Obviously,

$$\text{covar}(y, r_y) = \sigma_y \sigma_{r_y} \rho(y, r_y), \tag{2}$$

*Tel.: 001 202 473-6968; fax: 001 202 522-1153; e-mail: bmilanovic@worldbank.org

[1] See e.g. the original Corrado Gini definition as reported in Pyatt (1976), (p. 244) and discussed by Sen (1973), (p. 33); also Kakwani (1980); STATA software calculates the Gini in the same way; for matrix-based formulations, see Pyatt (1976); Berrebi and Silber (1987); Silber (1989); Milanovic (1994).

[2] Or more generally, recipients.

PII S0165-1765(97)00101-8

where σ_y = standard deviation of income, σ_{r_y} = standard deviation of individuals' ranks, and $\rho(y, r_y)$ = correlation coefficient between y and r_y. Now, writing out the standard deviation of ranks

$$\sigma_{r_y} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)^2}, \tag{3}$$

After some straightforward but tedious manipulation,

$$\begin{aligned}\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2 &= \sum_{i=1}^{N}\left(i^2 - i(N+1) + \frac{(N+1)^2}{4}\right) = \sum_{i=1}^{N} i^2 - (N+1)\sum_{i=1}^{N} i + \frac{1}{4}\sum_{i=1}^{N}(N+1)^2 \\ &= \frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{2} + \frac{N(N+1)^2}{4} \\ &= \frac{2N(N+1)(2N+1) - 3N(N+1)^2}{12} = \frac{4N^3 + 6N^2 + 2N - 3N^3 - 6N^2 - 3N}{12} \\ &= \frac{N^3 - N}{12} = \frac{N(N^2-1)}{12},\end{aligned}$$

Replacing the last expression into (3):

$$\sigma_{r_y} = \sqrt{\frac{N^2-1}{12}}. \tag{4}$$

Using (4), we can write:

$$G = \frac{2\,\mathrm{covar}(y, r_y)}{N\bar{y}} = \frac{2\sigma_y\sqrt{N^2-1}\,\rho(y, r_y)}{\sqrt{12}N\bar{y}} = \frac{1}{\sqrt{3}}\frac{\sigma_y}{\bar{y}}\rho(y, r_y)\frac{\sqrt{N^2-1}}{N}. \tag{5}$$

For a sufficiently large N, the last term in (5) will be equal to 1.[3] We thus obtain the final—simple—formula for the Gini coefficient:

$$G \approx \frac{1}{\sqrt{3}}\frac{\sigma_y}{\bar{y}}\rho(y, r_y). \tag{6}$$

which clearly shows that the Gini coefficient is the product of (i) a constant, (ii) coefficient of variation of income, and (iii) correlation coefficient between income and rank. Since all three elements are easy to calculate, the Gini can be obtained using a simple hand calculator. The concentration coefficient of a variable x when individuals are ranked according to y is straightforward:

$$G = \frac{1}{\sqrt{3}}\frac{\sigma_x}{\bar{x}}\rho(x, r_y).$$

[3]Note that even for $N = 10$ which may be thought the minimum sample size from which to calculate the Gini coefficient meaningfully the term amounts to 0.995. For $N = 30$, it is 0.99944.

1. The implications

Now, suppose that there is a linear relationship between income and rank. In other words, income always increases by the same amount as the ranks go up. Pen's parade (see Pen (1971), (p. 48ff)) is linear (see line A in Fig. 1). The linear correlation coefficient $\rho(y, r_y)$ must be equal to 1. We can write the linear relationship between income and rank as:

$$y = a + br_y,$$

Then,

$$\sigma_y = b\sigma_{r_y},$$

which using (4) yields

$$\sigma_y = b\frac{\sqrt{N^2-1}}{2\sqrt{3}}. \tag{7}$$

We also know that

$$\bar{y} = a + b\bar{r}_y = a + b\frac{N+1}{2}. \tag{8}$$

Combining (7) and (8), we obtain the coefficient of variation of y:

$$\frac{\sigma_y}{\bar{y}} = \frac{b\sqrt{N^2-1}}{2\sqrt{3}\left(a + b\frac{N+1}{2}\right)},$$

which after setting a to be arbitrarily small, simplifies

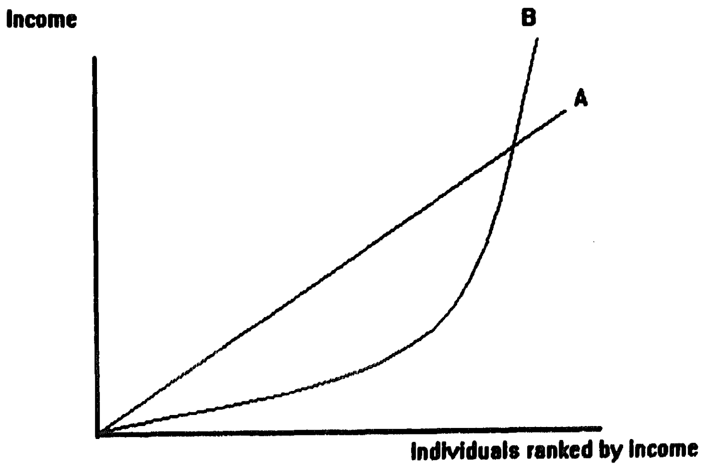

Fig. 1. Pen's parade.

$$\frac{\sigma_y}{\bar{y}}=\frac{b\sqrt{N^2-1}}{2\sqrt{3}\left(b\frac{N+1}{2}\right)}=\frac{\sqrt{N^2-1}}{\sqrt{3}(N+1)}=\frac{\sqrt{(N-1)(N+1)}}{\sqrt{3}(N+1)}=\frac{1}{\sqrt{3}}\sqrt{\frac{N-1}{N+1}}\approx\frac{1}{\sqrt{3}}. \tag{9}$$

Substituting (9) in (6) and remembering that in this case $\rho(y, r_y)=1$, we get:

$$G\approx\frac{1}{3}.$$

Thus, whenever Pen's parade is linear, the Gini coefficient must be, for a sufficiently large N, equal to $\frac{1}{3}$.[4,5]

However, we know that in almost all "real world" cases Pen's parade is convex (see line B in Fig. 1): incomes at first rise very slowly, and then their absolute increase, and finally even the rate of increase, accelerates. Thus, $\rho(y, r_y)$ in Eq. (6) which measures linear correlation will be less than 1.[6]

[4]The two approximations, from (5) and (9), combined yield:

$$\frac{\sqrt{N-1}}{\sqrt{N+1}}\frac{\sqrt{N^2-1}}{N}=\frac{N-1}{N},$$

which for $N=10$ gives 0.9 or for $N=100$, 0.99. Thus the true Gini of a linear Pen's parade with 10 observations will be $\frac{1}{3}$ times 0.9 = 0.3 (rather than $\frac{1}{3}$). But the true Gini of a linear parade with $N=100$ will be $\frac{1}{3}$ times 0.99, i.e. practically equal to $\frac{1}{3}$.

[5]This might, at first glance, seem odd, because linearity of Pen's parade holds for all kinds of b, such that the increases between successive individuals are small or large: e.g. if the distribution of five individuals is (1, 2, 3, 4, 5) or (100, 200, 300, 400, 500). Note, however, that these two distributions will —even if the absolute differences between the individuals are much larger in the second case—give the same Gini. This is not surprising because the relative (to the mean) income of each individual is the same in either case. And one distribution is simply a scalar transformation of the other distribution.

[6]Indeed, in all but two examples shown below and calculated from individual survey data, $\rho(y, r_y)$ ranges between 0.8 and 0.9. The countries are ranked by the coefficient of variation of income. Note a high negative correlation (−0.89) between the coefficient of variation and ρ.

Country (year)	N (individuals)	1 Coefficient of variation of income	2 Correlation coefficient $\rho(y, r_y)$	Gini coefficient $[1*2]/\sqrt{3}$
Hungary (1993; annual)	22 062	0.43	0.889	0.221
Poland (1993; annual data)	52 190	0.56	0.892	0.288
Romania (1994; monthly)	8999	0.57	0.863	0.284
Bulgaria (1994; annual)	7195	0.60	0.889	0.308
Estonia (1995; quarterly)	8759	0.68	0.871	0.342
UK (1986; annual)	71[illegible]8	0.68	0.815	0.320
Germany (1989; annual)	3940	0.71	0.744	0.305
US (1991; annual)	16 052	0.76	0.892	0.391
Russia (1993–4; quarterly)	16 356	1.07	0.812	0.502
Kyrghyzstan (1993; quarterly)	9547	1.63	0.586	0.551

Definition: recipients are individuals; income is disposable income per capita.

Source: Poland, Hungary and Estonia: household income surveys carried out by the countries' central statistical offices; Russia, Russia Longitudinal Monitoring Survey Round 3; Kyrghyz republic, Multipurpose Poverty Survey; Romania, Integrated household survey; Bulgaria, Gallup survey; UK, US and Germany from Luxembourg Income Study.

Now, the convex Pen parade will be obtained from a linear Pen parade through regressive transfers (from the poor to the rich). Such distribution, given by line B, must therefore be more unequal than the distribution given by line A. And, since $\rho(y, r_y)$ is now less than before, for the Gini to go up, the standard deviation of income must increase by more than ρ declines.[7]

The reverse of course holds for a concave Pen parade. A concave parade can be obtained by progressive transfers from the linear parade, and its Gini must be accordingly less than $\frac{1}{3}$. Both the standard deviation of income and ρ are decreasing.

We can thus conclude that all Gini coefficients of (around) $\frac{1}{3}$ are equivalent to the Gini that would obtain if Pen's parade were linear, that is if incomes were to increase by a constant absolute amount as ranks go up.[8] The Ginis below (over) $\frac{1}{3}$ are equivalent to the Ginis that would obtain if the parade were concave (convex).

2. Conclusions

In this note a simple way to calculate the Gini coefficient was derived. It was shown that the Gini coefficient is the product of three elements: (i) the coefficient of variation of the considered variable (say, income), (ii) income's linear correlation coefficients with ranks (1 to N), and (iii) a constant equal to $1/\sqrt{3}$. For all distributions such that the Pen's parade is linear, it was then shown that the Gini coefficient must be equal to $\frac{1}{3}$. For all distributions such that Pen's parade is concave (convex), the Gini coefficient must be smaller (greater) than $\frac{1}{3}$.

References

Berrebi, Z.M., Silber, J., 1987. Regional Differences and the Components of Growth and Inequality Change. Economic Letters 25, 295–298.

Kakwani, N., 1980. Income Inequality and Poverty: Methods of Estimation and Policy Analysis. World Bank and Oxford University Press, Washington and Oxford.

Lerman, R.I., Yitzhaki, S., 1984. A Note on the Calculation and Interpretation of the Gini Index. Economic Letters 15, 363–368.

Milanovic, B., 1994. The Gini-type Functions: An Alternative Derivation. Bulletin of Economic Research 1, 81–90.

Pen, J., 1971. Income Distribution: Facts, Theories, Policies. Praeger, New York.

Pyatt, G., 1976. On the Interpretation and Disaggregation of the Gini Coefficient. Economic Journal, June, 243–254.

Pyatt, G., Chen, C.-N., Fei, J., 1980. The Distribution of Income by Factor Components. Quarterly Journal of Economics, November, 451–473.

Sen, A., 1973. On Economic Inequality. Clarendon Press, Oxford.

Silber, J., 1989. Factor Components, Population Groups, and the Computation of the Gini Index of Inequality. Review of Economics and Statistics, 107–115.

Yitzhaki, S., 1994. Economic Distance and Overlapping of Distributions. Journal of Econometrics 61, 147–159.

[7] The mean is given.

[8] It does not mean that the Pen's parade underlying a Gini of $\frac{1}{3}$ is necessarily linear. The parade can be, for example, convex (with $\rho<1$) and such a coefficient of variation that their product (times $1/\sqrt{3}$) yields $\frac{1}{3}$.