the review of income and wealth

Review of Income and Wealth
Series 57, Number 3, September 2011
DOI: 10.1111/j.1475-4991.2011.00441.x

QUADRATIC PEN'S PARADE AND THE COMPUTATION OF THE GINI INDEX

BY STÉPHANE MUSSARD, J. SADEFO KAMDEM,* FRANÇOISE SEYTE AND MICHEL TERRAZA

Université Montpellier I

Following Milanovic's (1997) paper, we propose a simple way to compute the Gini index when income y is a quadratic function of its rank among n individuals.

1. Introduction

Suppose that positive incomes, expressed as a vector y, depend on individuals' ranks r_y in any given income distribution of size n. Suppose that incomes are ranked by ascending order and let $r_y = 1$ for the poorest individual and $r_y = n$ for the richest one. Hence, following Lerman and Yitzhaki (1984), the Gini index may be rewritten as follows:

$$G = \frac{2\operatorname{cov}(y, r_y)}{n\bar{y}}. \tag{1}$$

Here, $\operatorname{cov}(y, r_y)$ represents the covariance between incomes and ranks and $\bar{y}$ the mean income. It is straightforward to rewrite (1) as:

$$G = \frac{2\sigma_y \sigma_{r_y} \rho(y, r_y)}{n\bar{y}}, \tag{2}$$

where $\rho(y, r_y)$ is Pearson's correlation coefficient between incomes y and individuals' ranks r_y, where σ_y is the standard deviation of y, and where σ_{r_y} is the standard deviation of r_y.

Following (2) and under the assumption of a linear Pen's parade (i.e. $y = a + br_y$), Milanovic (1997) demonstrates that for a sufficiently large n, the Gini index can be further expressed as:

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

Milanovic's result is very interesting since it yields a simple way to compute the Gini index. However, as mentioned by Milanovic (1997, p. 48) himself, "in

*Correspondence to: Sadefo Kamdem, Université de Montpellier I, UFR Sciences Economiques, Avenue Raymond Dugrand, Site Richter, CS 79606, F-34960 Montpellier Cedex 2, France (sadefo@lameta.univ-montp1.fr).

almost all real world cases, Pen's parade is convex: incomes at first rise very slowly, and then their absolute increase, and finally even the rate of increase, accelerates." Thus, $\rho(y, r_y)$ which measures linear correlation will be less than 1. Again, from Milanovic (1997), a convex Pen's parade may be derived from a linear Pen's parade throughout regressive transfers (poor-to-rich income transfers). Inspired from Milanovic's finding, we demonstrate in the sequel, without taking recourse to regressive transfers, that the Gini index can be computed with a quite general quadratic Pen's parade.

2. Simple Gini Index with Quadratic Pen's Parade

Consider a quadratic relation between incomes and ranks:

$$y = a + br_y + cr_y^2. \tag{4}$$

The covariance between y and r_y is given by:

$$\operatorname{cov}(y, r_y) = b\operatorname{cov}(r_y, r_y) + c\operatorname{cov}(r_y^2, r_y) = b\sigma_{r_y}^2 + c\operatorname{cov}(r_y^2, r_y). \tag{5}$$

The mean income $\bar{y}$ is then:

$$\bar{y} = a + b\overline{r_y} + c\overline{r_y^2} = a + \frac{1}{n}\sum_{i=1}^{n}\left[bi + ci^2\right] = a + b\frac{n+1}{2} + c\frac{(n+1)(2n+1)}{6}. \tag{6}$$

2.1. *The Coefficient of Variation*

Since incomes y are positive, we use (1) by assuming that $c > 0$ and $b^2 - 4ac < 0$. We are now able to compute the coefficient of variation of incomes as follows:

$$\frac{\sigma_y}{\bar{y}} = \frac{\sqrt{b^2\sigma_{r_y}^2 + c^2\sigma_{r_y^2}^2 + 2bc\operatorname{cov}(r_y, r_y^2)}}{a + b\frac{(n+1)}{2} + c\frac{(n+1)(2n+1)}{6}}. \tag{7}$$

The standard deviation of r_y^2 is:

$$\begin{aligned} \sigma_{r_y^2} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} i^4 - \left(\frac{1}{n}\sum_{i=1}^{n} i^2\right)^2} \\ &= \sqrt{\frac{(n+1)(2n+1)(3n^2+3n-1)}{30} - \frac{(n+1)^2(2n+1)^2}{36}} \\ &= \frac{1}{6}\sqrt{\frac{(2n+1)(n+1)(8n^2+3n-11)}{5}}. \end{aligned} \tag{8}$$

The standard deviation of r_y is:

(9)
$$\sigma_{r_y} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} i\right)^2} = \sqrt{\frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}} = \sqrt{\frac{n^2-1}{12}}.$$

The covariance between r_y and r_y^2 is:

(10)
$$\operatorname{cov}\left(r_y, r_y^2\right) = \frac{1}{n}\sum_{i=1}^{n} i^3 - \left(\frac{1}{n}\sum_{i=1}^{n} i\right)\left(\frac{1}{n}\sum_{i=1}^{n} i^2\right) = \frac{n(n+1)^2}{4} - \frac{n+1}{2}\frac{(n+1)(2n+1)}{6} = \frac{(n+1)^2}{4}\frac{n-1}{3}.$$

Thereby, the coefficient of variation is expressed as:

(11)
$$\frac{\sigma_y}{\overline{y}} = \frac{\sqrt{\frac{b^2}{36}\frac{3\left(n^2-1\right)}{12} + \frac{c^2}{36}\frac{(2n+1)(n+1)\left(8n^2+3n-11\right)}{5} + 2bc\frac{n(n+1)^2}{2}\frac{n-1}{3}}}{a + b\frac{(n+1)}{2} + c\frac{(n+1)(2n+1)}{6}} = \frac{|c|}{c}\frac{\sqrt{\frac{b^2}{c^2}\frac{3\left(n^2-1\right)}{(2n+1)(n+1)} + \frac{8n^2+3n-11}{5(2n+1)(n+1)} + 12\frac{b}{c}\frac{n(n-1)}{(2n+1)^2}}}{\frac{6a}{c}\frac{1}{(2n+1)(n+1)} + \frac{6b}{c}\frac{1}{2n+1} + 1}.$$

Assuming that max $(a, b) \ll c$, since $c > 0$, we deduce the following limit:

(12)
$$\lim_{n\to\infty} \frac{\sigma_y}{\overline{y}} = \frac{2}{\sqrt{5}}.$$

Therefore, when n is sufficiently large, we have the following approximation for the coefficient of variation:

(13)
$$\frac{\sigma_y}{\overline{y}} \approx \frac{2}{\sqrt{5}}.$$

Remark 2.1 Under the quadratic Pen's parade assumption, we obtain an approximation for the coefficient of variation valued to be $2/\sqrt{5}$, while Milanovic (1997) obtained $1/\sqrt{3}$ under the linear Pen's parade assumption.

TABLE 1
Comparison between the Quadratic Gini Index and the Linear Gini Index

Country (year)	n	$\rho(y, r_y)$	$G = \frac{2}{\sqrt{15}}\rho(y, r_y)$	$G = \frac{1}{3}\rho(y, r_y)$
Hungary (1993; annual)	22,062	0.889	0.459	0.296
Poland (1993; annual)	52,190	0.892	0.461	0.297
Romania (1994; monthly)	8,999	0.863	0.446	0.288
Bulgaria (1994; annual)	7,195	0.889	0.459	0.296
Estonia (1995; quarterly)	8,759	0.871	0.450	0.290
U.K. (1986; annual)	7,178	0.815	0.421	0.272
Germany (1889; annual)	3,940	0.744	0.384	0.248
U.S. (1991; annual)	16,052	0.892	0.461	0.297
Russia (1993–4; quarterly))	16,356	0.812	0.419	0.271
Kyrgyzstan (1993; quarterly)	9,547	0.586	0.303	0.195

On the other hand, following Milanovic (1997):

$$\lim_{n\to\infty} 2\frac{\sigma_{r_y}}{n} = \lim_{n\to\infty}\sqrt{\frac{n^2-1}{3n^2}} = \frac{1}{\sqrt{3}}. \tag{14}$$

The product of (14), (13), and $\rho(y, r_y)$ entails the following result:

Theorem 2.1 Under the assumption of a quadratic Pen's parade, i.e., $y = a + br_y + cr_y^2$, the Gini index approximation is:

$$G \simeq \frac{2}{\sqrt{15}}\rho(y, r_y), \tag{15}$$

if n is sufficiently large, max $(a, b) \ll c$, $c > 0$ and $b^2 - 4ac < 0$.

2.2. *Application*

Following Milanovic's data (1997), we obtain the following results presented in Table 1.

Remark 2.2 As can be seen in Table 1, Milanovic's Gini index based on the linear Pen's parade underestimates the Gini index obtained under the quadratic Pen's parade.

3. Concluding Remarks

Following Milanovic (1997), we have proposed another simple way to calculate the Gini coefficient under the assumption of a quadratic Pen's parade.

Two immediate and practical implications result from this new Gini expression. First, the possibility to address a simplified signification test since our Gini index (as well as Milanovic's one) is based on Pearson's correlation coefficient. Thereby, testing for the Gini index signification reduces to testing for the signification of Pearson's correlation coefficient (up to the constant $2/\sqrt{15}$). This test relies on the well-known Student statistics based on the hyperbolic tangent transformation. Second, estimating the coefficients $\hat{a}$, $\hat{b}$, and $\hat{c}$, e.g. with Yitzhaki's Gini regression analysis, enables a parametric Gini index to be obtained that depends on parameters reflecting the curvature of Pen's parade, which may be of interest when one compares the shape of two income distributions.

References

Milanovic, B., "A Simple Way to Calculate the Gini Coefficient, and Some Implications," *Economics Letters*, 56, 45–9, 1997.

Lerman, R. I. and S. Yitzhaki, "A Note on the Calculation and Interpretation of the Gini Index," *Economics Letters*, 15, 363–8, 1984.