Gini Index and Polynomial Pen's Parade

Jules Sadefo Kamdem

Gini Index and Polynomial Pen's Parade

Jules SADEFO KAMDEM *
Université Montpellier I
Lameta CNRS UMR 5474 , FRANCE

Abstract

In this paper, we propose a simple way to compute the Gini index when income y is a finite order $k \in \mathbb{N}^*$ polynomial function of its rank among n individuals.

Key-words and phrases: Gini, Income inequality, Polynomial pen's parade, Ranks.

JEL Classification: D63, D31, C15.

*Corresponding author: Université de Montpellier I, UFR Sciences Economiques, Avenue Raymond Dugrand - Site Richter - CS 79606, F-34960 Montpellier Cedex 2, France; E-mail: sadefo@lameta.univ-montp1.fr

1 Introduction

Interest in the link between income and its rank is known in the income distribution literature as Pen's parade following Pen (1971, 1973). The precede has motivated research on the relationship between Pen's parade and the Gini index that is a very important inequality measure. However, this research has so far focused on the linear Pen's parade for which income increases by a constant amount as its rank increases by one unit (see Milanovic (1997). However, a linear parade does not closely fit many real world income distributions, Pen's parade, which is convex in the absence of negative incomes (i.e., incomes that increase by a greater amount as its rank increases by each additional unit). Mussard et al. (2010) has recently introduced the computation of Gini index with a convex quadratic Pen's parade (or second degree polynomial) parade for which income is a quadratic function of its rank.

In this paper, we extend the computation of gini index by using a more general and empirically more realistic case in which Pen's parade is a convex polynomial of finite order $k \in \mathbb{N}^*$. Hence, the Gini indexes for a linear Pen's parade and for quadratic pen's parade becomes a special case of that for an higher degree polynomial Pen's parade under some constraints on parameters to keep convexity.

The rest of the paper is organized as follows: In Section 2, the specification for a higher degree polynomial Pen's parade is provided. In Section 3, the problem of fitting a higher degree polynomial Pen's parade to real world data sets of Milanovic (1997) is discussed. Concluding is in Section 4.

2 High degree Polynomial Pen's Parade

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 in 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:

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

Here, $\text{cov}(y, r_y)$ represents the covariance between incomes and ranks and $\bar{y}$ the mean income. It is straightforward to rewrite (13) 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. $y = a + b\ r_y$), 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, page 48) himself, "in almost all real world cases, Pen's parade is convex: incomes rise very slowly at the beginning, then go up by their absolute increase, and finally increase even at the rate of acceleration". 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 by Milanovic's finding, we demonstrate in the sequel, without taking recourse to regressive transfers, that the Gini index can be computed with a general nonlinear polynomial function Pen's parade. The computation of the Gini Index using a polynomial function of order 2 (i.e. quadratic function) Pen's Parade has been introduced in Mussard et al.(2010). In this paper, we generalize the precedure to compute the Gini Index using a polynomial function of order $k \in \mathbb{N}^*$ pen's parade. Note that a linear pen's parade corresponds to $k = 1$ and quadratic pen's parade corresponds to $k = 2$. Our result here will be available for each finite $k \in \mathbb{N}^*$.

2.1 Simple Gini Index with nonlinear power Pen's Parade

Consider a power function relation between incomes and ranks:

$$y = \sum_{i=0}^{k-1} b_i\ r_y^i + b_k\ r_y^k\ . \tag{4}$$

with $k \in \mathbb{N}^*$. The covariance between y and r_y^j for $j \in \mathbb{N}^*$ is given by:

$$\begin{aligned} \operatorname{cov}(y, r_y^j) &= \sum_{i=0}^{k} b_i \operatorname{cov}\left(r_y^i, r_y^j \right) && (5) \\ &= \sum_{i=0}^{k} \left(\frac{b_i}{n} r_y^{i+j+1} - \frac{b_i}{n^2} \sum_{r_y=1}^{n} r_y^j \sum_{r_y=1}^{n} r_y^i \right) \\ &= \sum_{i=0}^{k} \left[\frac{b_i}{n} \sum_{r_y=1}^{n} r_y^{i+j+1} - b_i \sum_{r_y=1}^{n} r_y^i \sum_{r_y=1}^{n} r_y^j \right] \end{aligned}$$

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

$$\bar{y} = \sum_{i=0}^{k} b_i \, \overline{r_y^i} = \frac{1}{n} \left[\sum_{r=1}^{n} \sum_{i=0}^{k} b_i r^i \right] \qquad (6)$$

2.2 The coefficient of variation

Since the incomes y are positive, we use (13) by assuming that $b_k > 0$ and b_j for $j = 1, \ldots, k-1$ are chosen such that $y > 0$. For instance, if $k = 2$ then we can use b_j for $j = 1, \ldots, k$ such that $b_1^2 - 4b_2b_0 < 0$. We are now able to compute the coefficient of variation of incomes as follows:

$$\begin{aligned} \frac{\sigma_y}{\bar{y}} &= \frac{\sqrt{\sum_{i=0}^{k} \sum_{j=0}^{k} b_i b_j \operatorname{cov}(r_y^i, r_y^j)}}{\sum_{i=0}^{k} b_i \, \overline{r_y^i}} && (7) \\ &= \frac{\sqrt{\sum_{i=0}^{k} b_i \sum_{j=0}^{k} b_j \operatorname{cov}(r_y^i, r_y^j)}}{\sum_{i=0}^{k} b_i \, \overline{r_y^i}} && (8) \\ &= \frac{\sqrt{\sum_{j=0}^{k} b_j \operatorname{cov}(y, r_y^j)}}{\sum_{i=0}^{k} b_i \, \overline{r_y^i}} && (9) \\ &= \frac{\sqrt{\sum_{j=0}^{k} b_j \sum_{i=0}^{k} \left[\frac{b_i}{n} \sum_{r_y=1}^{n} r_y^{i+j+1} - \frac{b_i}{n^2} \sum_{r_y=1}^{n} r_y^i \sum_{r_y=1}^{n} r_y^j \right]}}{\frac{1}{n} \left[\sum_{r=1}^{n} \left(\sum_{i=0}^{k} b_i r^i \right) \right]}. \end{aligned}$$

where the variance of r_y^k is

$$\operatorname{cov}(r_y^k, r_y^k) = \sigma_{r_y^k}^2 = \frac{1}{n}\sum_{r=1}^{n} r^{2k} - \left(\frac{1}{n}\sum_{r=1}^{n} r^k\right)^2, \tag{10}$$

the covariance between r_y^j and r_y^i for $0 \leq i, j \leq k$, is:

$$\operatorname{cov}\left(r_y^i, r_y^j\right) = \frac{1}{n}\sum_{r=1}^{n} r^{i+j} - \frac{1}{n^2}\sum_{r=1}^{n} r^i \sum_{r=1}^{n} r^j. \tag{11}$$

After a double summation

$$\operatorname{cov}\left(\sum_{i=0}^{k} b_i r_y^i, \sum_{j=0}^{k} b_j r_y^j\right) = \sum_{i=0}^{k}\sum_{i=0}^{k} b_i b_j \left[\frac{1}{n}\sum_{r=1}^{n} r^{i+j} - \frac{1}{n^2}\sum_{r=1}^{n} r^i \sum_{r=1}^{n} r^j\right]. \tag{12}$$

Lemma 2.1 *When $n \to \infty$, for $q \in \mathbb{N}^*$ and $r \in \mathbb{N}^*$, we have that*

$$\sum_{r=1}^{n} r^q \equiv \frac{n^{q+1}}{q+1}. \tag{13}$$

Based on the preceding lemma, the variance of y when $n \to \infty$ is equivalent to:

$$\operatorname{cov}\left(\sum_{i=0}^{k} r_y^i, \sum_{j=0}^{k} r_y^j\right) \equiv \frac{b_k^2}{n}\frac{n^{k+k+1}}{k+k+1} - \frac{b_k^2}{n^2}\frac{n^{k+1}}{k+1}\frac{n^{k+1}}{k+1} \equiv \frac{(n^k b_k k)^2}{(2k+1)(k+1)^2}. \tag{14}$$

Therefore when $n \to \infty$ the standard deviation of y is equivalent to

$$\sigma_y \equiv \sqrt{\frac{(n^k b_k k)^2}{(2k+1)(k+1)^2}} \equiv \frac{k\,|b_k|}{(k+1)\sqrt{2k+1}} n^k. \tag{15}$$

When $n \to \infty$, the mean of y is equivalent to

$$\overline{y} = \frac{1}{n}\sum_{r=1}^{n}\sum_{i=0}^{k} b_i r^i \equiv \frac{b_k}{n}\frac{n^{k+1}}{k+1} \equiv b_k \frac{n^k}{k+1} \tag{16}$$

Thereby, as $n \to \infty$ the coefficient of variation is is equivalent expressed as:

$$\frac{\sigma_y}{\overline{y}} \equiv \frac{\frac{k\,|b_k|}{(k+1)\sqrt{2k+1}} n^k}{b_k \frac{n^k}{k+1}} \tag{17}$$

then we have the following limit which depends on k and the sign of b_k:

$$\lim_{n\to\infty} \frac{\sigma_y}{\overline{y}} = \frac{|b_k|}{b_k} \frac{k}{\sqrt{2k+1}} = sign(b_k)\frac{k}{\sqrt{2k+1}}. \tag{18}$$

We have then proved the following theorem:

Theorem 2.1 Under the assumption of a nonlinear polynomial Pen's parade, i.e., $y = \sum_{i=0}^{k} b_i\ r_y^i$, with $b_k \neq 0$, when $n \to \infty$, the coefficient of variation of the revenue y has the following limit:

$$\lim_{n\to\infty} \frac{\sigma_y}{\overline{y}} = \frac{|b_k|}{b_k} \frac{k}{\sqrt{2k+1}} \tag{19}$$

On the other hand, following Milanovic (1997):

$$\lim_{n\to\infty} 2\ \frac{\sigma_{r_y}}{n} = \lim_{n\to\infty} \sqrt{\frac{n^2-1}{3n^2}} = \frac{1}{\sqrt{3}}\ . \tag{20}$$

The product of (20), (18) and $\rho(y, r_y)$ entails the following result:

Theorem 2.2 Under the assumption of a nonlinear polynomial Pen's parade, i.e., $y = \sum_{i=0}^{k} b_i\ r_y^i$, with $b_k \neq 0$, when $n \to \infty$, the Gini index G_k can be approximately compute as follows:

$$G_k \simeq \frac{1}{\sqrt{3}} \frac{|b_k|}{b_k} \frac{k}{\sqrt{2k+1}} \rho(y, r_y), \tag{21}$$

where $\rho(y, r_y)$ is the correlation coefficient between the revenue and the rank r_y.

Remark 2.1 Note that for $k = 1$, we have the result of Milanovic (1997), i.e. $G_1 \simeq \frac{\rho(y,r_y)}{3}$ and for $k = 2$, we have the result of Sadefo Kamdem et al. (2010), i.e. $G_2 \simeq \frac{2\,\rho(y,r_y)}{\sqrt{15}}$.

Remark 2.2 Remarks that the Gini Index G_k approximation depends on k, the sign of b_k and $\rho(y, r_y)$. Since the Gini computation is independent of parameters b_i for $i = 0, \ldots, k-1$, we can compute the Pen's Parade as follows: $y = b_0 + b_k r^k$. Based on revenue data, it is simple to use a regression to estimate the parameters $\widehat{b_0}$ and $\widehat{b_k}$.

3 Application with Milanovic (1997) Data

In relation with the paramater k and $b_k > 0$, we have the following table:

Table 1: Computation of some coefficient of variation GV_k.

k	1	2	3	4	5	6	7	8
CV_k	0,577	0,894	1,134	1,333	1,508	1,664	1,807	1,940
GC_k	0.333	0.516	0.655	0.770	0.870	0.961	1.043	1.120

where CV_k denotes the coefficient of variation for polynomial Pen's Parade of order k and $GC_k = CV_k/\sqrt{3}$. Following Milanovic's data (1997), we obtain the following results:

Table 2: Comparison the gini indexes of G_k for $k = 1, 2, 3, 6$ with true G_{est}

Country (year)	n	$\rho(y, r_y)$	G_1	G_2	G_3	G_6	G_{est}
Hungary (1993; annual)	22062	0.889	**0.296**	0.459	0.582	0.854	**0.221**
Poland (1993; annual)	52190	0.892	**0.297**	0.461	0.584	0.857	**0.288**
Romania (1994; monthly)	8999	0.863	**0.288**	0.446	0.565	0.829	**0.284**
Bulgaria (1994; annual)	7195	0.889	**0.296**	0.459	0.582	0.854	**0.308**
Estonia(1995; quarterly)	8759	0.871	**0.290**	**0.450**	0.570	0.837	**0.342**
UK (1986; annual)	7178	0.815	**0.272**	**0.421**	0.534	0.783	**0.320**
Germany (1889; annual)	3940	0.744	0.248	**0.384**	0.487	0.715	**0.305**
US (1991; annual)	16052	0.892	0.297	**0.461**	0.589	0.857	**0.391**
Russia (1993-4; quarterly))	16356	0.812	0.271	0.419	**0.532**	0.780	**0.502**
Kyrgyzstan (1993; quarterly)	9547	0.586	0.195	0.303	0.384	**0.563**	**0.551**

In the preceding table, G_{est} denotes the estimation of the true Gini index by using Milanovic data (1997).

Remark 3.1 Based on the analysis of the preceding table, we propose to consider linear Pen's parade (k=1) to compute the Gini index of Poland (1993, annual), Romania (1994, monthly), Bulgaria(1994; annual). For estonia (1995; quarterly) and UK (1986; annual), we can also choose $k = 1$, but we consider $k = 2$ in the case where government policy prefers to overestimate inequality instead of underestimate inequality. For Russia (1993-4; quarterly), we consider $k = 3$ and for Kyrgyzstan (1993; quarterly) we choose $k = 6$.

Remark 3.2 In our polynomial pen's parade, if $r_y = 1$, then $y = y_1 = \sum_{i=1}^{k} b_k$. In practical applications, the parameters $\widehat{b}_k$, estimated from the observed data using multiple regression, are chosen so that the revenue $y > 0$ (i.e. $b_k > 0$). It's very important to note that, the correlation coefficient between y and its rank r_y depend on b_k. Ignoring this fact will arbitrarily restrict the Parade to pass through the origin and may result in less accurate estimates of the Gini Index.

4 Concluding Remarks

Following Milanovic (1997), we have proposed another simple way to calculate the Gini coefficient under the assumption of a general polynomial Pen's Parade of order k . By using the data in Table 2, we conclude that the computation of the Gini Index of each revenues data needs to find a specific integer k, which is the order of a specific polynomial Pen's Parade. Our Gini computation is useful in practical applications as soon as the limit expression obtained in (21)is a good approximation of the Gini Index for usual size n.

Two immediate and practical applications can be generated from this new Gini expression. First, the possibility to address a significant test since our Gini Index (as well as Milanovic's) is based on Pearson's correlation coefficient. Thereby, testing for the Gini Index significance is equivalent to testing for the significance of Pearson's correlation coefficient (up to the constant $\frac{k}{\sqrt{2k+1}}$). This test relies on the well-known student statistics based on the hyperbolic tangent transformation. Second, estimating the coefficients b_i for $i = 1, \ldots, k$, e.g. with Yitzhaki's Gini regression analysis or a multiple regression, enables a parametric Gini Index to be obtained. That procedure depends on parameters reflecting the curvature of Pen's Parade. This may be of interest when one compares the shape of two income distributions.

References

[1] Milanovic, B., "A simple way to calculate the Gini coefficient, and some implications," *Economics Letters* , 56, 45-49, 1997.

[2] Lerman R.I., S. Yitzhaki, "A note on the calculation and interpretation of the Gini index," *Economics Letters*, 15, 363-368, 1984.

[3] Pen, J., Income Distribution. *(Allen Lane The Penguin Press, London)*, 1971.

[4] Pen, J., A parade of dwarfs (and a few giants),*in: A.B. Atkinson, ed., Wealth, Income and Inequality: Selected Readings, (Penguin Books, Middlesex)* 73-82, 1973.

[5] Mussard, S., Sadefo Kamdem, J., Seyte, F., Terraza, M., "Quadratic Pen's parade and the computation of Gini Index, *Accepted for publication in Review of Income and Wealth*, 2010.