

Coefficient of variation and Power Pen's parade computation

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

April 16, 2011

Abstract

Under the the assumption that income y is a power function of its rank among n individuals, we approximate the coefficient of variation and gini index as functions of the power degree of the Pen's parade. Reciprocally, for a given coefficient of variation or gini index, we propose the analytic expression of the degree of the power Pen's parade; we can then compute the Pen's parade.

Key-words and phrases: Gini index, Income inequality, Ranks, Harmonic Number, Pen's Parade.

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-montpl.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 Pens parade and the Gini index that is a very important inequality measure. However, this research has so far focused on the linear Pens parade for which income increases by a constant amount as its rank increases by one unit (see Milanovic (1997). Furthermore, a linear pen's parade does not closely fit many real world income distributions Pens parade of which is convex in the absence of negative incomes (i.e., income increases 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 convex quadratic Pens parade (or second degree polynomial) parade for which income is a quadratic function of its rank (see also Ogwang (2010)).

This paper extends the computation of gini index by using a more general and empirically more realistic case for which Pens parade is a power function. Hence, the Gini indices for a linear Pen's parade and for quadratic Pen's parade becomes special cases of our thus introduced power Pens parade under some constraints on parameters inducing convexity. Under the assumption that income y is a power function of its rank among n individuals, we approximate the coefficient of variation and gini index as functions of the power degree of the pen's parade. Reciprocally, for a given coefficient

of variation or gini index, we propose the analytic expression of the degree of the power pen's parade. It follows that, for a given coefficient of variation or gini index we can compute the pen's parade. We use the Milanovic (1997) data to illustrate the interest of our results.

The rest of the paper is organized as follows: In Section 2, the specification of an a power Pens parade is provided. In Section 3., the problem of fitting a power Pens parade to real world datasets of Milanovic (1997) is discussed. The concluding remarks are made in Section 4.

2 High degree Power 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 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:

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

Here, $\operatorname{cov}(y, r_y)$ represents the covariance between incomes and ranks and \bar{y} the mean income. It is straightforward to rewrite (13) as:

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

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:

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

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 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 nonlinear polynomial function Pen's parade. The computation of the gini index using a power 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 precede to compute gini index using a power function pen's parade. Note that a linear and quadratic Pen's

parade are particular cases our power Pen parade.

2.1 Simple Gini Index with nonlinear power Pen's Parade

Consider a power function relation between incomes and ranks:

$$y = \sum_{i=0}^{k-1} b_i r^{\alpha_i} + b_k r^{\alpha_k} . \quad (4)$$

with $k \in \mathbb{N}^*$, $\alpha_0 = 0$ and $\alpha_i \in \mathbb{R}_+^*$ for $i = 1, \dots, k$, and

$$\alpha_k > \max\{\alpha_i, i = 1, \dots, k-1\}.$$

The covariance between y and r^{α_j} for $j \in \mathbb{N}^*$ is given by:

$$\begin{aligned} \text{cov}(y, r^{\alpha_j}) &= \sum_{i=1}^k b_i \text{cov}(r^{\alpha_i}, r^{\alpha_j}) \\ &= \sum_{i=1}^k \left(\frac{b_i}{n} r^{\alpha_i + \alpha_j + 1} - \frac{b_i}{n^2} \sum_{r=1}^n r^{\alpha_i} \sum_{r=1}^n r^{\alpha_j} \right) \\ &= \sum_{i=1}^k \left[\frac{b_i}{n} \sum_{r=1}^n r^{\alpha_i + \alpha_j + 1} - \frac{b_i}{n^2} \sum_{r=1}^n r^{\alpha_i} \sum_{r=1}^n r^{\alpha_j} \right] \end{aligned} \quad (5)$$

The mean income \bar{y} is then:

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

2.2 The coefficient of variation and Gini computation

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

$$\frac{\sigma_y}{\bar{y}} = \frac{\sqrt{\sum_{i=1}^k \sum_{j=1}^k b_i b_j \operatorname{cov}(r^{\alpha_i}, r^{\alpha_j})}}{\sum_{i=1}^k b_i \bar{r}^{\alpha_i}} \quad (7)$$

$$= \frac{\sqrt{\sum_{i=1}^k b_i \sum_{j=1}^k b_j \operatorname{cov}(r^{\alpha_i}, r^{\alpha_j})}}{\sum_{i=1}^k b_i \bar{r}^{\alpha_i}} \quad (8)$$

$$= \frac{\sqrt{\sum_{j=1}^k b_j \operatorname{cov}(y, r_y^j)}}{\sum_{i=1}^k b_i \bar{r}^{\alpha_i}} \quad (9)$$

$$= \frac{\sqrt{\sum_{j=1}^k b_j \sum_{i=1}^k \left[\frac{b_i}{n} \sum_{r=1}^n r_y^{\alpha_i + \alpha_j + 1} - \frac{b_i}{n^2} \sum_{r_y=1}^n r^{\alpha_i} \sum_{r=1}^n r^{\alpha_j} \right]}}{\frac{1}{n} \left[\sum_{r=1}^n \left(\sum_{i=0}^k b_i r^{\alpha_i} \right) \right]}.$$

where the variance of r^{α_k} is

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

the covariance between r^{α_j} and r^{α_i} for $0 < i, j \leq k$, is:

$$\operatorname{cov}(r^{\alpha_i}, r^{\alpha_j}) = \frac{1}{n} \sum_{r=1}^n r^{\alpha_i + \alpha_j} - \frac{1}{n^2} \sum_{r=1}^n r^{\alpha_i} \sum_{r=1}^n r^{\alpha_j}. \quad (11)$$

After a double summation

$$\text{cov} \left(\sum_{i=1}^k b_i r^{\alpha_i}, \sum_{j=1}^k b_j r^{\alpha_j} \right) = \sum_{i=1}^k \sum_{i=1}^k b_i b_j \left[\frac{1}{n} \sum_{r=1}^n r^{\alpha_i + \alpha_j} - \frac{1}{n^2} \sum_{r=1}^n r^i \sum_{r=1}^n r^{\alpha_j} \right]. \quad (12)$$

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

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

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

$$\text{cov} \left(\sum_{i=1}^k r^{\alpha_i}, \sum_{j=1}^k r^{\alpha_j} \right) \equiv \frac{b_k^2}{n} \frac{n^{\alpha_k + \alpha_k + 1}}{\alpha_k + \alpha_k + 1} - \frac{b_k^2}{n^2} \frac{n^{\alpha_k + 1}}{\alpha_k + 1} \frac{n^{\alpha_k + 1}}{\alpha_k + 1} \equiv \frac{(n^{\alpha_k} b_k k)^2}{(2\alpha_k + 1)(\alpha_k + 1)^2}. \quad (14)$$

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

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

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

$$\bar{y} = \frac{1}{n} \sum_{r=1}^n \sum_{i=0}^k b_i r^{\alpha_i} \equiv \frac{b_k}{n} \frac{n^{\alpha_k + 1}}{\alpha_k + 1} \equiv b_k \frac{n^{\alpha_k}}{\alpha_k + 1} \quad (16)$$

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

$$\frac{\sigma_y}{\bar{y}} \equiv \frac{\frac{\alpha_k |b_k|}{(\alpha_k+1)\sqrt{2\alpha_k+1}} n^{\alpha_k}}{b_k \frac{n^{\alpha_k}}{\alpha_k+1}} \quad (17)$$

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

$$\lim_{n \rightarrow \infty} \frac{\sigma_y}{\bar{y}} = \frac{|b_k|}{b_k} \frac{\alpha_k}{\sqrt{2\alpha_k+1}} = \text{sign}(b_k) \frac{\alpha_k}{\sqrt{2\alpha_k+1}}. \quad (18)$$

We have then proved the following theorem:

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

$$\lim_{n \rightarrow \infty} \frac{\sigma_y}{\bar{y}} = \frac{|b_k|}{b_k} \frac{\alpha_k}{\sqrt{2\alpha_k+1}} \quad (19)$$

On the other hand, following Milanovic (1997):

$$\lim_{n \rightarrow \infty} 2 \frac{\sigma_{r_y}}{n} = \lim_{n \rightarrow \infty} \sqrt{\frac{n^2-1}{3n^2}} = \frac{1}{\sqrt{3}}. \quad (20)$$

Remark 2.1 The following theorem provide an interesting result in practise. It tells us that, for a given income data such as the coefficient of variation, we can compute the degree of the power Pen's parade. Therefore, we can compute the Pen's parade.

Theorem 2.2 For $b_k > 0$, If we know the coefficient of variation CV_{emp} by using incomes data, then we can find the parameter $s = \alpha_k$ as a solution of the following equation:

$$\frac{s}{\sqrt{2s+1}} = CV_{emp}. \quad (21)$$

We can then use multiple regression to find b_i and get the Pen's parade that corresponds to incomes data. After a straightforward calculus, the positive solution of the equation (25) is :

$$\alpha_k = CV_{emp} \left[CV_{emp} + \sqrt{CV_{emp}^2 + 1} \right]. \quad (22)$$

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

Theorem 2.3 Under the assumption of a power Pen's parade, i.e., $y = \sum_{i=0}^k b_i r^{\alpha_i}$, with $b_k \neq 0$, when $n \rightarrow \infty$, the Gini index G_k can be approximately computed as follows:

$$G_{\alpha_k} \simeq \frac{1}{\sqrt{3}} \frac{|b_k|}{b_k} \frac{\alpha_k}{\sqrt{2\alpha_k+1}} \rho(y, r_y), \quad (23)$$

where $\rho(y, r_y)$ is correlation coefficient between the incomes y and their ranks r_y .

Corollary 2.1 For $b_k > 0$, the Gini index is approximately equal to

$$G_{\alpha_k} \simeq \frac{1}{\sqrt{3}} \frac{|b_k|}{b_k} \frac{\alpha_k}{\sqrt{2\alpha_k+1}} \rho(y, r_y), \quad (24)$$

where α_k is the solution of the following equation:

$$\frac{s}{\sqrt{2s+1}} = CV_{emp}. \quad (25)$$

where CV_{emp} is the coefficient of variation obtained by using incomes data y .

Remark 2.2 Note that for $\alpha_k = 1$, we have the result of Milanovic (1997), i.e. $G_1 \simeq \frac{\rho(y, r_y)}{3}$ and for $\alpha_k = 2$, we have the result of Mussard et al. (2010), i.e. $G_2 \simeq \frac{2\rho(y, r_y)}{\sqrt{15}}$.

Remark 2.3 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 do not depends on b_i for $i = 0, \dots, k-1$, we can then compute the Pen's parade as follows: $y = b_0 + b_k r^{\alpha_k}$. Based on revenues data, it is convenient to use a simple regression to estimate the parameters \hat{b}_0 and \hat{b}_k . Also we can use the following simplify power Pen's parade $y = b_0 + b_1 r + b_k r^{\alpha_k}$. If the revenues are ordered as follows: $0 \leq y_1 \leq y_2 \leq \dots \leq y_n$, then the precede Pen's parade pass through the origin $(1, y_1)$ implies that $y_1 = b_0 + b_1 + b_k$, $y_2 = b_0 + 2b_1 + 2^{\alpha_k} b_k$ and $y_3 = b_0 + 3b_1 + 3^{\alpha_k} b_k$ where α_k is given by 33. We can the find b_0 , b_1 and b_k .

2.3 Correlation between the incomes and ranks

The covariance between income y and the rank r_y is

$$\begin{aligned}
 cov(y, r_y) &= \sum_{i=0}^k b_i [cov(r^{\alpha_i}, r)] \\
 &= \sum_{i=1}^k b_i \left[\frac{1}{n} \sum_{r=1}^n r^{\alpha_i+1} - \frac{1}{n^2} \sum_{r=1}^n r^{\alpha_i} \sum_{r=1}^n r \right] \\
 &= \sum_{i=1}^k b_i \left[\frac{1}{n} H[n, -\alpha_i - 1] - \frac{n+1}{n} H[n, -\alpha_i] \right] \quad (26)
 \end{aligned}$$

where

$$H[n, s] = \sum_{r=1}^n \frac{1}{r^s}$$

is an harmonic number function for $s > 1$ and $n \in \mathbb{N}^*$. In the simple case where, $y = b_0 + b_1 r^{\alpha_1}$, we have:

$$cov(y, r_y) = b_1 \left[\frac{1}{n} H[n, -\alpha_1 - 1] - \frac{n+1}{n} H[n, -\alpha_1] \right] \quad (27)$$

The variance of the income y is:

$$cov(y, y) = b_1^2 \left[\frac{1}{n} H[n, -\alpha_1 - 1] - \left(\frac{1}{n} \sum_{r=1}^n H[n, -\alpha_1] \right)^2 \right]. \quad (28)$$

The standard deviation of income is:

$$\sigma_y = \sqrt{cov(y, y)} = |b_1| \sqrt{\frac{1}{n} H[n, -\alpha_1 - 1] - (H[n, -\alpha_1])^2}. \quad (29)$$

The standard deviation deviation of individuals ranks is given by:

$$\sigma_{r_y} = \sqrt{\frac{1}{n} \sum_{r=1}^n r^2 - \left(\frac{1}{n} \sum_{r=1}^n r \right)^2} = \frac{(2n+1)(n+1)}{6} - \frac{(n+1)^2}{4} \quad (30)$$

The correlation coefficient between y and r is:

$$\begin{aligned} \rho(y, r_y) &= \frac{b_1 \left[\frac{1}{n} \sum_{r=1}^n r^{\alpha_1+1} - \frac{n+1}{n} \sum_{r=1}^n r^{\alpha_1} \right]}{|b_1| \sqrt{\frac{1}{n} \sum_{r=1}^n r^{\alpha_1+1} - \left(\frac{1}{n} \sum_{r=1}^n r^{\alpha_1} \right)^2} \sqrt{\frac{n^2-1}{12}}} \\ &= \frac{b_1 \left[\frac{1}{n} H[n, -\alpha_1 - 1] - \frac{n+1}{n} H[n, -\alpha_1] \right]}{|b_1| \sqrt{\frac{1}{n} H[n, -\alpha_1 - 1] - (H[n, -\alpha_1])^2} \sqrt{\frac{n^2-1}{12}}}. \end{aligned} \quad (31)$$

Theorem 2.4 If the income $y = b_0 + b_k r^{\alpha_k}$, with $\alpha_1 > 0$ and $b_k \neq 0$ (i.e. $b_k > 0$), then the Gini index is approximately equal to

$$G_{\alpha_k} \simeq \frac{1}{\sqrt{3}} \frac{\alpha_k}{\sqrt{2\alpha_k + 1}} \frac{\frac{1}{n} H[n, -\alpha_k - 1] - \frac{n+1}{n} H[n, -\alpha_k]}{\sqrt{\frac{1}{n} H[n, -\alpha_k - 1] - (H[n, -\alpha_k])^2} \sqrt{\frac{n^2-1}{12}}} \quad (32)$$

where

$$\alpha_k = CV_{emp} \left(CV_{emp} + \sqrt{1 + CV_{emp}} \right) \quad (33)$$

with CV_{emp} being the coefficient of variation obtained by using income data y .

3 Pen's parade of different countries data and their Gini indexes

In this section, we use the Milanovic (1997) countries incomes data for the computation of the Pen's parade and gini index of each country income data. In relation with the parameter k and $b_k > 0$, we have the following table:

Table 1: The Positive solution $s = \alpha_k$ (see (22)) of equation (25) in relation with the empirical coefficient of variation CV_{emp} of each country.

CV_{emp}	0.43	0.56	0.57	0.60	0.68	0.68	0.71	0.76	1.07	1.63
α_k	0.653	0.955	0.981	1.060	1.285	1.285	1.375	1.532	2.712	5.774

Table 2: Coefficient of variation and Gini index G_{α_k} for different degree α_k for each country Pen's parade.

Country (year)	n	$\rho(y, r_y)$	CV_{emp}	α_k	G_{α_k}
Hungary (1993; annual)	22062	0.889	0.43	0.653	0.221
Poland (1993; annual)	52190	0.892	0.56	0.955	0.288
Romania (1993; annual)	52190	0.892	0.57	0.981	0.288
Bulgaria (1994; monthly)	8999	0.863	0.60	1.060	0.284
Estonia(1995; quarterly)	7195	0.889	0.68	1.285	0.308
UK (1986; annual)	8759	0.871	0.68	1.285	0.342
Germany (1889; annual)	7178	0.815	0.71	1.375	0.320
US (1991; annual)	3940	0.744	0.76	1.532	0.305
Russia (1993-4; quarterly))	16052	0.892	1.07	2.712	0.391
Kyrgyzstan (1993; quarterly)	16356	0.812	1.63	5.774	0.502

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

Remark 3.1 Based on the analysis of the previous table, we can propose to consider power Pen's parade ($\alpha_k = 0.653$) to compute the Gini index for Hungary (1993, annual), $\alpha_k = 0.955$ for Poland(1993; annual), $\alpha_k = 0.981$ for Romania (1994, monthly), $\alpha_k = 1.060$ for Bulgaria(1994; monthly), $\alpha_k = 1.285$ for Estonia (1995; quarterly), $\alpha_k = 1.285$ for UK (1986; annual), $\alpha_k = 1.375$ for Germany (1889; annual), $\alpha_k = 1.532$ for US (1991; annual),

$\alpha_k = 2.712$ for Russia (1993-4; quarterly)) and $\alpha_k = 5.774$ for Kyrgyzstan (1993; quarterly).

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

4 Concluding Remarks

In this paper, we proposed the power pen's parade as an alternative to the linear and quadratic pen's parade. Following Milanovic (1997), we have proposed another simple way to calculate the Gini coefficient under the assumption of a general power Pen's parade of order k . By using the data of Table 2, we concluded that the computation of the Gini index of each income data country need to find a specific $\alpha_k \in \mathbb{R}_+^*$ which is the order of a specific power Pen's parade. We have therefore compute the Pen's parade and the gini index for each of the six countries.

It appeared that for a given coefficient of variation of a country incomes data, we found the associated degree of the power Pen's parade, hence enabling compute the pen's parade using multiple regression. It also provides

a convenient and straightforward way to compute the Gini index of each country by using the coefficient of correlation between incomes and ranks.

One immediate and practical implications result from this new Gini expression. Estimating the coefficients b_i for $i = 1, \dots, k$, e.g. with Yitzhaki's Gini regression analysis or a multiple regression, 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

- [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.
- [6] Ogwang, T., "The gini index for a quadratic prn's parade, *Working paper*, 2010.