Gini coefficient estimation using parabolic model

Xiaoqi Niu[1]*, Baoping Chan [2]

[1] *School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, Henan, China*

[2] *School of software engineering, Anyang Normal University, Anyang 455000, Henan, China*

Received 1 June 2014, www.cmnt.lv

Abstract

The goal of this paper is to establish a parabolic model for Gini Coefficient estimation. With the help of computer simulation and programming, this paper develops a new method of Gini Coefficient estimation and provides statistic support for the administrative and the decision-making departments of the government to scientifically measure the income gaps and to adjust the distribution policy and macroeconomic policy.

Keywords: Gini Coefficient, Lorenz Curve, Parabola Model, Coefficient Estimation

1 Introduction

The Gini coefficient which developed by Corrado Gini in1912 is a measure of statistical dispersion intended to represent the income distribution of a nation's residents, and is the most commonly used measure of inequality. Its range is [0, 1], and the lower the value the distribution of wealth among social members more uniform. And the 0.4 is usually used as the income distribution gap between the warning lines [1]. Even a little mistake of one percentage in mathematics can cause a great error which may mislead the economists and decision-making departments of the government into misjudgement of the gaps in the income distribution and may further influence the adjustment of income distribution policies and macro-economic policies. Therefore, it is essential to choose a suitable algorithm with relatively smaller error based on the data collected.

The Gini coefficient is widely used. The literature [2] shows explicitly how the overlapping of groups impacts between-group inequality by generalizing a result on the group-wise decomposition of the Gini index to more than two groups. The literature [3] shows the ranking of income distributions, symmetric under the same transformation, by S-Gini consistent social evaluation functions and majority voting coincide if and only if the inequality index under consideration is the Gini coefficient. The spatial Lorenz Curve and Gini coefficient can be used to analysis the Gini Coefficient of agricultural production, revealing how the agricultural specialization changed and analyzing the status of a certain city in the agricultural specialization of the whole province [4]. The Gini coefficient is also can be used to design the company salary system in order to improve the staff's work enthusiasm and reduce their turnover rate [5]. Estimation of Gini Coefficient is an approximate calculation [6-8]. There have been many traditional methods, such as the trapezoidal method, the rectangle method, and the Gini calculation method. However, the Simpson method or the parabola method, which plays an important role in approximate calculation, has never been mentioned in the literatures.

The purpose and innovation point of this paper is to propose a new mathematical model for estimation of Gini Coefficient: the parabola method. Moreover, computer programming of the mathematical model is developed to perform automation of Gini Coefficient estimation so as to offer statistical support to the administrative and decision-making departments of the government to make scientific judgment of the gaps in income distribution and right adjustments of distribution policy and macro-economic policy.

2 Parabolic Model for Gini Coefficient Estimation

The trapezoidal method and the rectangle method [9, 10] divide Lorenz Curve into several sections and replace the curve with straight line in every section. In other words, they replace the Lorenz Curve with linear function. When the Lorenz Curve has a relatively large degree of curvature, the error in Gini Coefficient estimation may be enlarged and will obviously underestimate the Gini Coefficient.

In order to improve the accuracy, it is practical to replace Lorenz Curve with the quadratic function of $I = ap^2 + bp + c$ in a small range. Exactly, we replace a section of the original Lorenz Curve with the corresponding section of a parabolic curve, whose symmetric axis is parallel to the y axis. This method is called Parabolic Method.

Lemma 1 Three non-collinear points can determine only one parabola.

Lemma 2 The parabola $y = ax^2 + bx + c$ over $[x_0, x_0+h]$ which is determined by three non-collinear points: (x_0,

* ***Corresponding author's*** e-mail: nxq@aynu.edu.cn

y_0), (x_0+h/2, y_1), (x_0+h, y_2) can form a curve-edge trapezoid with the area of

$$s=\int_{x_0}^{x_0+h}(ax^2+bx+c)dx=\frac{1}{6}h(y_0+4y_1+y_2) \tag{1}$$

The area of such a trapezoid is only related with the y coordinates of the three points respectively and the length of the interval mentioned above.

To further generalize the above conclusion: let $0=x_0<x_1<x_2<...<x_n=1$ be a partition of [0, 1] which divides [0, 1] into n intervals (n is an even number) with equal length. The length of every interval is $1/n$, and the function value corresponding to every division point is y_0, y_1, y_2 ,…, y_n respectively. The curve of $y=f(x)$ is also divided into n parts with division points M_0, M_1, M_2,…, M_n respectively. As a result, the parabolas determined respecttively by three non-collinear points of (M_0, M_1, M_2; M_2, M_3, M_4; …; M_{n-2} M_{n-1} M_n) can form different curve-edge trapezoids with the area respectively of :

$$S_{\frac{n}{2}}=\frac{1}{3n}(y_{n-2}+4y_{n-1}+y_n) \tag{2}$$

Where n is an even number. If we add together the areas of the above n/2 trapezoids, we can get the total area as

$$\int_0^1 f(x)dx\approx\frac{1}{3n}[(y_0+y_n)+2(y_2+y_4+\cdots+y_{n-2}) +4(y_1+y_3+\cdots+y_{n-1})] \tag{3}$$

The above formula is the Parabolic Formula for definite integral approximate calculation, or known as Simpson Formula.

From the above mathematical expression, we can get the formula of parabolic method for Gini coefficient calculation.

Theorem: If we divide the total population into n groups from the lowest income to the highest income (n must be an even number), the population of every group is 1/n of the total population, let I_1, I_2, I_3,…,I_n denote the proportion of income of every group in the total income of all members respectively, then

$$G=1-\frac{2}{3n}[1+2(y_2+y_4+\cdots+y_{n-2})+4(y_1+y_3+\cdots+y_{n-1})] \tag{4}$$

Where

$$y_i=\sum_{k=1}^{i}I_k \quad (i=1,2,...,n) \tag{5}$$

Based on the proof [9], we know $G=1-2S_B$ where S_B can be calculated with Formula (3). Note that $y_0=0, y_n=1$, then we can get Formula (4).

3 Estimation flow of Parabolic Methods for Gini Coefficient

We use parabolic method to estimate the Gini coefficient. The estimation flows please see Figure 1.

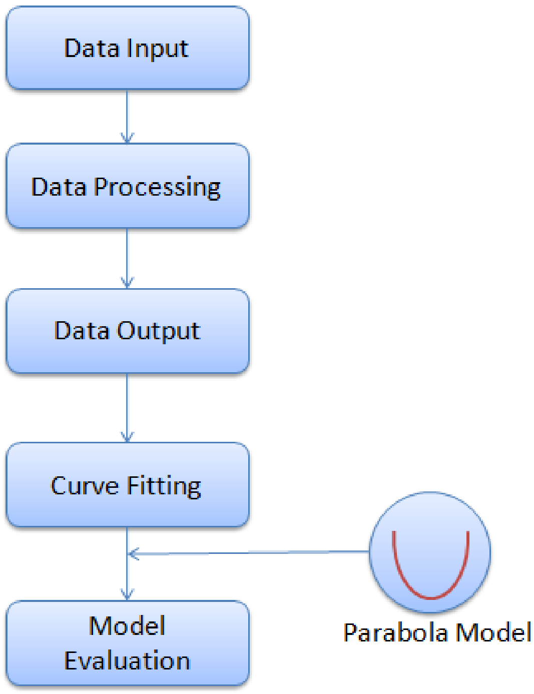

FIGURE 1 Gini coefficient estimation flow.

In Figure 1, the data will be inputted by using user interface. The stage of data processing is a preparation for Gini Coefficient calculations. The output data are discrete points which returned from the stage of data processing. As output data given here are discrete, it is impossible to draw the Lorenz Curve precisely. The Lorenz Curve drawn here is formed by connecting the scattered points. At last, the fitted curve will be evaluated by using the parabolic model.

4 Examples and Analysis

4.1 IDEAL SITUATION OF INCOME DISTRIBUTION

Suppose there are n members in the sample space, and the income of the lowest one is a units of some currency, and the difference between two adjacent members is one unit. Obviously, this is an ideal structure of income distribution, so the total income of all the members is:

$$\begin{aligned}A&=\sum_{i=1}^{n}A_i\\&=a+(a+1)+(a+2)+\cdots+[a+(n-2)]+[a+(n-1)]\\&=\frac{n}{2}[2a+(n-1)]\end{aligned} \tag{6}$$

Among all the members, the rate of income of i th member compared with the total income of all members is

$$I_i=\frac{A_i}{A}(1\le i\le n).$$

Then we can list all these n numbers according to their incomes from the lowest to the highest. Please see Table 1.

TABLE 1 Sorted incomes of the n numbers

No.	Income (A_i)	Per. of Population	Per. of Income	Cumulative Per. of Population	Cumulative Per. of Incomes
1	a	$\frac{1}{n}$	$\frac{a}{A}$	$\frac{1}{n}$	$\frac{a}{A}$
2	a+1	$\frac{1}{n}$	$\frac{a+1}{A}$	$\frac{2}{n}$	$\frac{2a+1}{A}$
3	a+2	$\frac{1}{n}$	$\frac{a+2}{A}$	$\frac{3}{n}$	$\frac{3(a+1)}{A}$
…	…	…	…	…	…
n-1	a+(n-2)	$\frac{1}{n}$	$\frac{a+(n-2)}{A}$	$\frac{n-1}{n}$	$\frac{n-1}{2A}[2a+(n-2)]$
n	a+(n-1)	$\frac{1}{n}$	$\frac{a+(n-1)}{A}$	1	1

4.2 GINI COEFFICIENT IN IDEAL SITUATION OF INCOME DISTRIBUTION

The Gini coefficient calculation formula with Parabolic Method is as follows:

$$G = 1 - \frac{2}{3n}[1 + 2(y_2 + y_4 + \cdots + y_{n-2}) + 4(y_1 + y_3 + \cdots + y_{n-1})]$$

$$= 1 - \frac{2}{3n}\{1 + 2[\frac{2a+1}{A} + \frac{4a+6}{A} + \cdots + \frac{n-2}{2A}(2a+n-3)]$$

$$+4[\frac{a}{A} + \frac{3(a+1)}{A} + \cdots + \frac{n-1}{2A}(2a+n-2)]\} \quad (7)$$

$$= 1 - \frac{2}{3n}\{1 + \frac{2}{A}\sum_{k=1}^{\frac{n}{2}-1}\frac{2k}{2}[2a+(2k-1)] + \frac{4}{A}\sum_{k=1}^{\frac{n}{2}}\frac{2k-1}{2}[2a+(2k-2)]\}$$

$$= 1 - \frac{2}{3n}\left\{1 + \frac{2}{A}\sum_{k=1}^{\frac{n}{2}-1}\left[(2a-1)k + 2k^2\right] + \frac{4}{A}\sum_{k=1}^{\frac{n}{2}}\left[(1-a) + (2a-3)k + 2k^2\right]\right\}$$

where $$A = \frac{n}{2}\left[2a + (n-1)\right] \quad (8)$$

We can use

$$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1) \quad (9)$$

to refine the formula:

$$G = 1 - \frac{2}{3n}\{1 + \frac{2n}{A}(1-a) + \frac{(n-2)n}{4A}(2a-1) + \frac{1}{6A}(n-2)(n-1)n$$
$$+ \frac{n(n+2)}{2A}(2a-3) + \frac{1}{3A}n(n+1)(n+2)\} \quad (10)$$

At last, we can get the result, the optimal value of Gini coefficient is 1/3 theoretically. See Equation (11).

$$\lim_{n\to\infty} G = 1 - \frac{2}{3}(\frac{1}{3} + \frac{2}{3}) = \frac{1}{3} \quad (11)$$

If $a=1$, $n=100$, $A=5050$, and put the values of P_i, I_i into the estimation flow of Gini coefficient calculation, the value of Gini coefficient got by the Parabolic method is 0.33, It is close to the warning value 0.4. The income distribution curve using the Gini coefficient discussed above is shown as Figure 2.

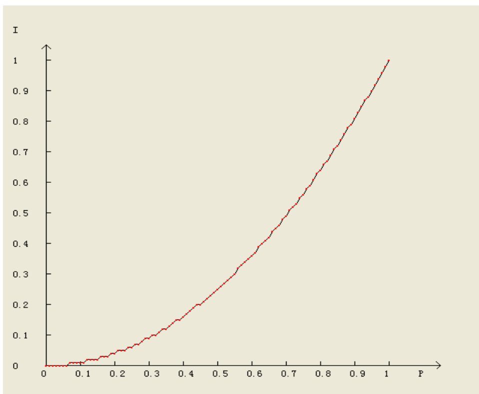

FIGURE 2. Income distribution curve using Gini coefficient.

As we can see from Figure 2, the experimental curve is almost a perfect parabola. So it is feasible to forecast the income using our model design and implement.

5 Conclusions

The first contribution is that it has introduced a new mathematic model for Gini coefficient estimation-Parabola Model. When the Lorenz Curve has a relatively large degree of curvature, this Parabola Model has a higher accuracy than other models. The second contribution is to realize the automation of the Gini Coefficient Parabola Model estimation.

Acknowledgments

This work was supported by the Soft science project of Henan Province (132400410153).

References

[1] Zhao Guangxi (2013) Research on the Trend and Countermeasures for the Change of Gini Coefficient in China. *Journal of Finance and Economics Theory*, (6): 39-43.
[2] Allanson P (2014) Income stratification and between-group inequality. *Economics letters*. 124(2), 227-230.
[3] Rodriguez J G, Salas R. (2014) The Gini coefficient: Majority voting and social welfare. *Journal of economic theory*, 152(1): 214-223.
[4] Peng Xiangyu, Tao Sxiqi, Guo Yanxin (2014) Discussion on Specialization of Agricultural Production in Hubei Province from the Research Perspective of Jingzhou City Based on Spatial Lorenz Curve and Gini Coefficient. *Modern Agricultural Science and Technology*, (1): 310-312.
[5] Lu Wei (2014) Using Gini coefficient design enterprise salary system. *Human Resource Management*. (1): 90-91.
[6] Li Hongyi, Xie Danyang, and Zou Hengfu (2000) Dynamics of income Distribution. *Canadian Journal of Economics*, 4: 937-961
[7] F. A. Cowell (1998) *Prepared for Handbook of Income Distribution*. Edited by A. B. Atkinson and F. Bourguignon.
[8] Barro R (2000) Inequality and Growth in a Panel of Countries. *Journal of Economic Growth*, 65: 5-32.
[9] Niu Xiaoqi (2009) A Study to Prove Equivalence of Three Methods of Gini Coefficient Calculation. *Journal of Xinyang Normal University*, (3): 364-367.
[10] Xiaoqi N (2009) Three Key Indexes Related with Gini Coefficient. *Journal of Henan Normal University*, (5): 26-28.

Authors

Xiaoqi Niu, 25. 2. 1963, China

Xiaoqi Niu received the master degree in department of mathematics of Beijing Normal University, China in 1991. Currently, he is an associate professor at Anyang Normal University, China. His research interests include Mathematical modeling and calculation method, and he has published more than 20 papers in these fields.

Baoping Chang, 13. 11. 1964, China

Baoping Chang received the Bachelor degree in department of mathematics of Beijing Normal University, China in 1986. Currently, he is a professor at School of software engineering, Anyang Normal University, China. His research interests include computer application technology, computer network and application, and he has published more than 20 papers in these fields.