Research & Reviews: Journal of of Statistics and Mathematical Sciences

Estimating Lorenz Curve of Income in China by Cubic Spline Interpolation

Ying Liang[1], Jiaxu Zhang[2]*, Dandan Jia[2], Kungui Zhang[2], Qiang Li[2] and Xuezhang Liang[2]

[1]Business School, Jilin University, Changchun, Jilin, 130012, P.R. China

[2]Mathematics School, Jilin University, Changchun, Jilin, 130012, P.R. China

RESEARCH ARTICLE

Received date: 29/08/2015

Accepted date: 20/11/2015

Published date: 30/11/2015

***For Correspondence**

Jiaxu Zhang, Mathematics School, Jilin University, Changchun, Jilin, 130012, P.R. China

E-mail: liangxz@jlu.edu.cn

Keywords: Cubic spline interpolation, Lorenz curve, Gini index, Lorenz curve aggregation.

ABSTRACT

In this paper, we use a kind of cubic-spline-interpolation to approximate the Lorenz curves of urban and rural area in China, respectively. According to the Gini index reported by State Statistics Bureau of China in January 2014, we aggregate the Lorenz curves of urban and rural area in China, and obtain the national Lorenz curves of whole China from 2003 to 2013.

INTRODUCTION

Lorenz curve and Gini index are important tools to measure the income distribution of the residents. Normally Lorenz curve is defined on the interval[0,1]and it is increasing, convex, continuous curve. Else, Lorenz curve satisfies $L(0)=0, L(1)=1$. As the **Figure 1** shows, the curve $OE_1E_2E_3E_4L$ is the Lorenz curve, the diagonal OL means income distribution is totally equal. On the other hand, the broken line OXL means the income is distributed nonuniform extremely. The bigger distance between Lorenz and diagonal, the more unequal income distribution is. A lot of papers have been published in this área [1-5]. The Lorenz curves of rural and urban area from 2003 to 2013 have been reported in China Statistical Yearbook, but the Lorenz curve of whole China has not published yet. In this paper, we firstly use a kind of cubic spline interpolation [6] to approximate Lorenz curves of urban and rural area in this paper, respectively. Then we propose an aggregation approach to obtain the Lorenz curve of whole China referring to the Gini index (2003-2013) reported by State Statistics Bureau in January 2014. We select the average income of rural area to average income of urban area ratio α appropriately, and obtain the national Lorenz curve and Gini index using our proposed aggregation approach, such that the Gini index of each year equals to the Gini index reported by the State Statistics Bureau of China exactly.

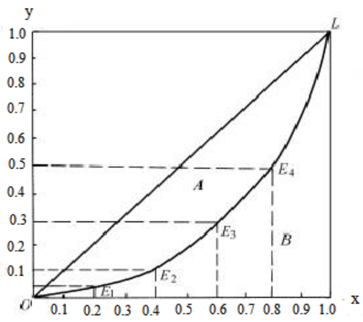

Figure 1. Lorenz curve.

A kind of cubic spline interpolation

In this paper, we use a kind of cubic-spline-interpolation at a special endpoint condition(6) to estimate Lorenz curve. In this way, the advantages of the cubic spline interpolation are that it can fit grouped data exactly and maintain the convexity of the curve in most cases. The procedure is as following:

The Lorenz curve is a monotonously increasing convex function **(Figure 1)**, this is means that the first derivative of the Lorenz curve is greater than zero . and second derivative of the Lorenz curve is non-negative. We approximate the Lorenz curve by a cubic spline interpolation function $y = s(x)$.

Denote the interpolation knots by $\{x_i\}_{i=0}^{n}$, ($0 \le x_i \le 1, (i = 0,1,2 \ldots n)$), the corresponding values of the cubic spline function $s(x)$ in the knots are $\{y_i\}_{i=0}^{n}$, ($0 \le y_i \le 1, (i = 0,1,2 \ldots n)$). The second derivative $s''(x)$ of $s(x)$ in each interval is linear function. Let $m_i = s''(x_i)$. We have

$$s''(x) = m_{i-1}\frac{x_i - x}{h_i} + m_i\frac{x - x_{i-1}}{h_i}\ , \quad \forall x \in [x_{i-1}, x_i]\ , \quad \text{where} \quad h_i = x_i - x_{i-1}\ , \quad i = 1,2 \ldots n\ . \tag{1}$$

While $x \in [x_i, x_{i+1}]$,

$$s\ (x) = m_{i-1}\frac{(x_i - x)^3}{6h_i} + m_{i+1}\frac{(x - x_{i-1})^3}{6h_i} + (y_{i-1} - \frac{m_{i-1}h_i^2}{6})\frac{x_i - x}{h_i} + (y_i - \frac{m_i h_i^2}{6})\frac{x - x_{i-1}}{h_i} \tag{2}$$

Here , m_i satisfies the following equations:

$$\frac{h_i}{6}m_{i-1} + \frac{h_i + h_{i+1}}{3}m_i + \frac{h_{i+1}}{6}m_{i+1} = \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}\ , i = 1,2,3 \ldots n-1\ . \tag{3}$$

with two boundary conditions:

$$\begin{cases} 2m_0 + \lambda_0 m_1 = d_0 \\ \mu_n m_{n-1} + 2m_n = d_n \end{cases} \quad \text{and} \quad \begin{cases} \lambda_i = \dfrac{h_{i+1}}{h_i + h_{i+1}} \\ \mu_i = 1 - \lambda_i \end{cases} \tag{4}$$

Usually people take the boundary conditions as following:

$$\begin{cases} d_0 = \dfrac{6}{h_1}(\dfrac{y_1 - y_0}{h_1} - y_0') \\ d_n = \dfrac{6}{h_n}(y_n' - \dfrac{x_n - x_{n-1}}{h_n}) \\ \lambda_0 = \mu_n = 1 \end{cases} \tag{5}$$

But under the boundary conditions(5)above, the cubic spline curve may not satisfy the convexity requirement.

Through the numerical experiment, we found when taking

$$\begin{cases} \lambda_0 = \mu_n = -2 \\ d_0 = d_n = 0 \end{cases} \tag{6}$$

the approximated Lorenz curves we tested are always monotonously increasing and smooth convex function. This point is the key of our approach.

In this condition the formula of Gini index is

$$G = 1 - 2B = 1 - \sum_{i=0}^{n-1}\frac{y_i + y_{i+1}}{h_i} + \frac{1}{12}\sum_{i=0}^{n-1}\frac{m_i + m_{i+1}}{h_i^3} \tag{7}$$

A lorenz curve aggregation formula

Let P_1 be the population of urban area, and P_2 be the population of rural area, $P = P_1 + P_2$ denote the national population. The population of urban area to nation ratio is $\lambda_1 = \frac{P_1}{P}$,for rural area is $\lambda_2 = \frac{P_2}{P}$.

Income distribution function and it's inversion :

$$p_i = F_i(x)\ , \ x_i = F_i^{-1}(p_i) = x_i(p_i) \tag{8}$$

where p_i is cumulative population share, x is income

Density function of income distribution :

$$\frac{dp_i}{dx} = F_i'(x) = f_i(x) \tag{9}$$

Lorenz curve:

$$L_i(p_i) = \frac{1}{\mu_i}\int_0^{p_i} x_i(q)dq = \frac{1}{\mu_i}\int_0^{x} tf_i(t)dt \tag{10}$$

Average income μ_i:

$$\mu_i = \int_0^{+\infty} xf_i(x)dx \tag{11}$$

Derivation of Lorenz curve:

$$\frac{dL_i}{dp_i} = L_i'(p_i) = \frac{1}{\mu_i}x_i(p_i) \tag{12}$$

$i = 1,2$ (while i=1, it denotes urban area i=2, it denotes rural area:

Aggregation income distribution function:

$$p = \lambda_1 p_1 + \lambda_2 p_2 = \lambda_1 F_1(x) + \lambda_2 F_2(x) = F(x) \tag{13}$$

Our derivation of Aggregation Lorenz curve is as following :

$$L = L(p) = \frac{1}{\mu}\int_0^{p} x(q)dq \quad = \frac{1}{\mu}\int_0^{x} tf(t)dt \quad = \frac{1}{\mu}\int_0^{x} t[\lambda_1 f_1(t) + \lambda_2 f_2(t)]$$

$$= \frac{1}{\mu}[\lambda_1\int_0^{x} tf_1(t)dt + \lambda_2\int_0^{x} tf_2(t)dt]$$

$$= \frac{1}{\mu}[\lambda_1\mu_1 L_1(p_1) + \lambda_2\mu_2 L_2(p_2)] \tag{14}$$

Where aggregation average income:

$$\mu = \lambda_1\mu_1 + \lambda_2\mu_2 \tag{15}$$

Obviously we have

$$x = L_1'(p_1)\mu_1 = L_2'(p_2)\mu_2 = L'(p)\mu \tag{16}$$

Our aggregation formula is as following:

$$\begin{cases} p = \lambda_1 p_1 + \lambda_2 p_2 \\ x = L_1'(p_1)\mu_1 = L_2'(p_2)\mu_2 \\ L(p) = \dfrac{1}{\mu}[\lambda_1\mu_1 L_1(p_1) + \lambda_2\mu_2 L_2(p_2)] \end{cases} \tag{17}$$

Estimating lorenz curve of China by aggregation approach

Firstly, according to the data from China Statistical Yearbook (2014) in [7], we do the numerical computations by cubic spline interpolation formula (3),(4) and (6), and obtain the Lorenz curve and Gini index of rural and urban area.(2003-2013). Then we estimate the Lorenz curve of whole China by our aggregation formula.

From the formula (17) we can see that when $\lambda_1, \lambda_2, L_1(p_1)\ L_2(p_2)$ are given, the aggregating Lorenz curve $L(p)$ and the Gini index only depend on the ratio $\alpha = \frac{\mu_2}{\mu_1}$ of each year (the values of μ_1 and μ_2 can magnify or reduce at this ratio). Comparing the Gini index reported in State Statistics Bureau in January 2014, we adjust the value of α and calculate the aggregation Lorenz curves and Gini indexies (2003-2013), such that the calculation Gini index of each year equals to Gini index reported in State Statistics Bureau exactly. Finally we get the national Lorenz curve of China (2003-2013). The computation results are as following **Tables 1 and 2.**

Table 1. The values of and rural, urban, nationwide Gini indexies of China (2003-2013).

Year	Rural Gini Index	Urban Gini Index	α value	Nationwide Gini index
2003	0.3694	0.3185	0.2631	0.479
2004	0.3584	0.3255	0.2696	0.473
2005	0.3647	0.3309	0.2528	0.485
2006	0.3627	0.3272	0.2440	0.487
2007	0.3629	0.3243	0.2421	0.484
2008	0.3668	0.3305	0.2338	0.491
2009	0.3742	0.3256	0.2280	0.490
2010	0.3675	0.3200	0.2323	0.481
2011	0.3766	0.3193	0.2351	0.477
2012	0.3753	0.3058	0.2224	0.474
2013	0.3729	0.3046	0.2162	0.473

Table 2. The values and values of Chinese lorenz curve form 2003-2013 of our proposed approach.

	α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
2003	0.2631	0.0132	0.0366	0.0700	0.1110	0.1687	0.2463	0.3518	0.4850	0.6614	0.8021
2004	0.2696	0.0139	0.0384	0.0730	0.1152	0.1752	0.2512	0.3557	0.4870	0.6614	0.8025
2005	0.2528	0.0126	0.0355	0.0683	0.1086	0.1684	0.2426	0.3455	0.4768	0.6554	0.8002
2006	0.2440	0.0122	0.0346	0.0664	0.1064	0.1629	0.2419	0.3441	0.4766	0.6572	0.8024
2007	0.2421	0.0118	0.0341	0.0659	0.1061	0.1636	0.2436	0.3464	0.4804	0.6618	0.8064
2008	0.2338	0.0111	0.0325	0.0634	0.1031	0.1599	0.2389	0.3403	0.4747	0.6590	0.8055
2009	0.2280	0.0103	0.0309	0.0609	0.1009	0.1589	0.2389	0.3413	0.4780	0.6639	0.8095
2010	0.2323	0.0109	0.0321	0.0627	0.1045	0.1641	0.2457	0.3501	0.4874	0.6720	0.8150
2011	0.2351	0.0100	0.0313	0.0623	0.1061	0.1671	0.2501	0.3555	0.4912	0.6756	0.8176
2012	0.2224	0.0097	0.0300	0.0597	0.1040	0.1666	0.2503	0.3587	0.4987	0.6846	0.8241
2013	0.2162	0.0094	0.0298	0.0589	0.1038	0.1667	0.2504	0.3606	0.4994	0.6858	0.8250

CONCLUSION

In this paper, we proposed a kind of cubic spline interpolation to estimate Lorenz curve and an approach to aggregate the Lorenz curve of urban and rural areas. Through the numerical experiment, we apply theses tow approaches to the income data of rural an urban of China and get the aggregation Lorenz curves of China (2003-2013). The computation results show that these two approaches are effective. The results we get have some reference value.

ACKNOWLEDGEMENT

The work is supported by grant (11271041) of Chinese National Science Foundation and Jilin University Philosophy and Social Science Fund (2011QY093).

REFERENCES

1. Dagum C. A New Approach to the Decomposition of the Gini Income Inequality Ratio Empirical Economics 1997; 22: 515-531
2. Chotikapanich D et al. Estimating Income Inequality in China Using Grouped Data and the Generalized Beta Distribution. Review of Income and Wealth 2007; 53: 127-47.
3. Wang ZX et al. A General Method for Creating Lorenz Curves. Reviews of Income and Wealth 2011; 3: 561-582.
4. Huang W. Numerical Method to Calculate Gini Coefficient from Limited Data of Subgroups. Applied Economics Letters 2013; 20: 1249-1253
5. Liang Y, et al. A method of estimating Gini index from grouped data. 2014 International Conference on Humanity an Social Science 2014; 255-258.
6. De Boor C. A Practical Guide to Splines.New York. Springer-Verlig. 1978.
7. National Bureau of China. China Statistical Yearbook. China Statistics Press 2014.