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 $OE, E, E, L$ 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 area [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 China, and obtain the national Lorenz curves of whole China from 2003 to 2013.
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 \( \{ x_i \}_{i=0}^n \), \( 0 \leq x_i \leq 1, (i = 0, 1, 2, \ldots, n) \), the corresponding values of the cubic spline function \( s(x) \) in the knots are \( \{ y_i \}_{i=0}^n \), \( 0 \leq y_i \leq 1, (i = 0, 1, 2, \ldots, n) \). The second derivative \( s''(x) \) of \( s(x) \) in each interval is linear function.

Let \( m_i = s'(x) \). We have

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

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

\[
 s(x) = m_{i-1} \left( x - x_i \right)^3 \frac{6h_i}{6h_i} + m_i \left( x - x_{i+1} \right)^3 \frac{6h_{i+1}}{6h_{i+1}} + \left( y_{i-1} - m_{i-1} \frac{h_i^2}{6} \right) \frac{x - x_i}{h_i} + \left( y_i - m_i \frac{h_{i+1}^2}{6} \right) \frac{x - x_{i+1}}{h_{i+1}}
\]

Here, \( m_i \) satisfies the following equations:

\[
 \begin{align*}
 &2m_b + \lambda_0 m_1 = d_0 \\
 &\mu_a m_{n+1} + 2m_n = d_n
\end{align*}
\]

\[ \lambda_i = \frac{h_{i+1}}{h_i + h_{i+1}} \]

(3)

with two boundary conditions:

\[
 \begin{align*}
 &d_0 = \frac{6}{h_i} (y_i - y_0 - y'_0) \\
 &d_n = \frac{6}{h_{i+1}} (y_n' - x_{n-1} - h_{i+1}) \\
 &\lambda_0 = \mu_n = 1
\end{align*}
\]

(4)

Usually people take the boundary conditions as following:

\[
 \begin{align*}
 &d_0 = \frac{6}{h_i} (y_i - y_0 - y'_0) \\
 &d_n = \frac{6}{h_{i+1}} (y_n' - x_{n-1} - h_{i+1}) \\
 &\lambda_0 = \mu_n = 1
\end{align*}
\]

(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

\[
 \begin{align*}
 &\lambda_0 = \mu_n = -2 \\
 &d_0 = d_n = 0
\end{align*}
\]

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

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

(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 \( \lambda_0 = \frac{P_1}{P} \) for rural area is \( \lambda_2 = \frac{P_2}{P} \).

Income distribution function and it’s inversion:

\[
 p_i = F_i(x), \quad x_i = F_i^{-1} (p_i) = x_i (p_i)
\]

where \( p_i \) is cumulative population share, \( x_i \) is income

Density function of income distribution:
\( \frac{dp_i}{dx} = F_i(x) = f_i(x) \) \quad (9)

Lorenz curve:
\[
L_i(p_i) = \frac{1}{\mu_i} \int_0^p x_i(q) dq = \frac{1}{\mu_i} \int_0^p tf_i(t) dt
\]
\quad (10)

Average income \( \mu_i \):
\[
\mu_i = \int_0^{\infty} xf_i(x) dx
\]
\quad (11)

Derivation of Lorenz curve:
\[
\frac{dL_i}{dp_i} = L'_i(p_i) = \frac{1}{\mu_i} x_i(p_i)
\]
\quad (12)

\( i = 1, 2 \) (while \( i = 1 \), it denotes urban area; \( i = 2 \), it denotes rural area):

Aggregation income distribution function:
\[
p = \lambda_1 p_1 + \lambda_2 p_2 = \lambda_1 F_i(x) + \lambda_2 F_2(x) = F(x)
\]
\quad (13)

Our derivation of Aggregation Lorenz curve is as following
\[
L = L(p) = \frac{1}{\mu} \int_0^p x(q) dq = \frac{1}{\mu} \int_0^p tf(t) dt = \frac{1}{\mu} \int_0^p [\lambda_1 f_1(t) + \lambda_2 f_2(t)]
\]
\[
= \frac{1}{\mu} [\lambda_1 \int_0^p t f_1(t) dt + \lambda_2 \int_0^p t f_2(t) dt]
\]
\[
= \frac{1}{\mu} [\lambda_1 \mu_1 L_1(p_1) + \lambda_2 \mu_2 L_2(p_2)]
\]
\quad (14)

Where aggregation average income:
\[
\mu = \lambda_1 \mu_1 + \lambda_2 \mu_2
\]
\quad (15)

Obviously we have
\[
x = L'_i(p_i) \mu_i = L'_2(p_2) \mu_2 = L'(p) \mu
\]
\quad (16)

Our aggregation formula is as following:
\[
\begin{cases}
\lambda_1 p_1 + \lambda_2 p_2 \\
x = L'_i(p_i) \mu_i = L'_2(p_2) \mu_2 \\
L(p) = \frac{1}{\mu} [\lambda_1 \mu_1 L_1(p_1) + \lambda_2 \mu_2 L_2(p_2)]
\end{cases}
\]
\quad (17)

**Estimating Lorenz curve of China by aggregation approach**

Firstly, according to the data from China Statistical Yearbook (2014) \(^{[7]}\), 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 \( \mu_1 \) and \( \mu_2 \) can magnify or reduce at this ratio). Comparing the Gini index reported in State Statistics Bureau in January 2014, we adjust the value of \( \alpha \) and calculate the aggregation Lorenz curves and Gini indexes (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 α and rural, urban, nationwide Gini indexes of China (2003-2013).

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

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

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

### 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 these two approaches to the income data of urban and rural areas 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