Analytical Examination of GINI Index and Lorenz Curves: A Scientific Inquiry

Lingfei Kong

University of Rochester, Rochester, NY,14627, United States

Keywords: Gini Index; Lorenz Curve; Curve-fitting Method; Riemann Sum Method

Abstract: The Gini Index, derived from the Lorenz curve, serves as a benchmark for assessing wealth inequality within a nation. Employing calculus and curve-fitting techniques to establish the Lorenz curve model not only furnishes a theoretical foundation for more effectively tracking the income distribution among residents but also holds practical implications. Additionally, a thorough examination of the Lorenz curve's first and second derivatives enables a profound analysis of how alterations in national income and population dynamics impact the curve, offering valuable insights into the intricacies of income distribution patterns.

1. Background

In economics, the Gini index is a vital measure of people's well-being. This paper explores the challenges in studying the Gini index and Lorenz curve, focusing on the following questions:

Recent World Bank studies reveal social unrest in Latin American countries, despite their higher average income compared to many Southeast Asian nations. The key factor is the Gini index, reflecting the rich-poor gap. Latin American countries exhibit a high Gini index of 0.522, surpassing their economic development levels, linking this to persistent social unrest.

To address societal tensions, economists and mathematicians have delved into Gini index research, emphasizing precise calculation methods. This paper employs various approaches, comparing and optimizing accurate calculation methods. Additionally, examining Lorenz curve derivatives aims to deeply analyze how changes in national income and population impact curve dynamics.

This comprehensive approach seeks to enhance our understanding of income distribution, aiding informed policy decisions to address social disparities[1].

2. Problem statement

In economics, the Gini index plays a crucial role in gauging people's life satisfaction. As we delve into the complexities of studying the Gini index and Lorenz curve, various challenges emerge. This paper focuses on addressing the following inquiries:

- Assessing National Income Disparities and Happiness: How can we effectively measure the overall disparity in a country's national income and subsequently evaluate the national happiness index?
- Minimizing Errors in Gini Index Estimation: What methods can be employed to minimize errors when estimating the Gini Index, ensuring a more accurate reflection of wealth distribution?
- Analyzing Lorentz Curve Dynamics: How do we analyze the rate of change of the Lorentz curve and understand the significance of its changing patterns?

This paper aims to employ calculus function models to systematically quantify and address these issues, contributing to a more scientific understanding of the complexities associated with wealth distribution and its impact on overall well-being.

3. Brief introduction about the Gini Index and Lorenz Curve

3.1 Lorenz Curve

American statistician Lorenz lists the cumulative percentage of the country's total population as a

  DOI: 10.25236/iceesr.2024.022

horizontal coordinate, ranks the total population of the country from low to high, and lists the cumulative percentage of income formed as a vertical coordinate, and then plots the correspondence between the cumulative percentage of the total population and the cumulative percentage of income on a graph, which is the Lorenz curve (See Figure 1).

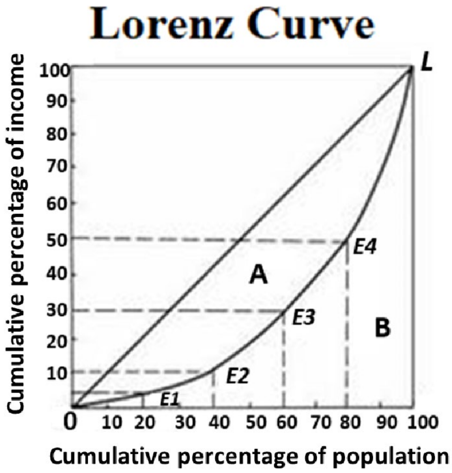

Figure 1. Lorenz curve

The curvature of the Lorenz curve holds significant meaning, serving as an indicator of income distribution inequality. A more pronounced bend signifies greater inequality, while a gentler curve suggests a more even distribution. Notably, complete income inequality results in a 90-degree polyline when all income concentrates in one person's hands, while perfect equality forms a 45-degree line through the origin when every percentage of the population equals its income percentage. Understanding the Lorenz curve provides insights into the nuances of income distribution dynamics, enabling a comprehensive assessment of economic equality within a population[2-3].

3.2 The Gini Index

The Gini coefficient is a widely utilized measure globally for assessing the income disparity among residents in a country or region. In the context of the Lorentz curve, the Gini index is calculated as the ratio of area A to the sum of areas A and B.

Key points regarding the Gini coefficient include:

The Gini coefficient ranges from a minimum of 0, indicating absolute income equality, to a maximum of 1, representing absolute income inequality.

The actual Gini coefficient falls within the 0 to 1 range.

A Gini coefficient below 0.19 indicates a highly even income distribution.

For coefficients between 0.19-0.25, the income distribution is considered relatively average.

In the 0.25-0.40 range, the income distribution is characterized as basically average.

A Gini coefficient surpassing 0.40 signifies a highly uneven income distribution.

Understanding the Gini coefficient provides valuable insights into the degree of income inequality within a given population, facilitating nuanced assessments of economic equity[4-5].

4. Mathematical model for the estimation of the Gini Index

4.1 Curve-fitting method

Precise calculation of the Gini Index hinges on accurately determining the area A of the tension in the Lorentz curve. This study employs various functional curve-fitting methods to calculate area A and scrutinize its impact on the Gini index.

The table below, derived from data provided by the US Census Bureau, illustrates Lorenz function values for income distribution in the United States for the year 2008. This dataset serves as a foundation for the subsequent analysis (See Table 1):

Table 1. Lorentz Function Values

X	0.0	0.2	0.4	0.6	0.8	1.0
L(x)	0.000	0.034	0.120	0.267	0.500	1.000

- Curve-fitting Function Analysis: $L(x) = Ax^2+Bx$

We begin by assuming $L(x) = Ax^2+Bx$ and evaluating its impact on the calculation of area A in the Lorentz curve.

First Iteration: Assuming L(x) = 0.65x2+0.04x. Applying data points (x=0.2, L(x)=0.034)and (x=0.4, L(x) = 0.12) yields the function. Further points x=0.6, 0.8 and 1.0 are incorporated, resulting in L(x) = 0.258, 0.448 and 0.69. When graphed, the calculated area A exceeds the actual value. Leading to an inflated GINI index (See Figure 2).

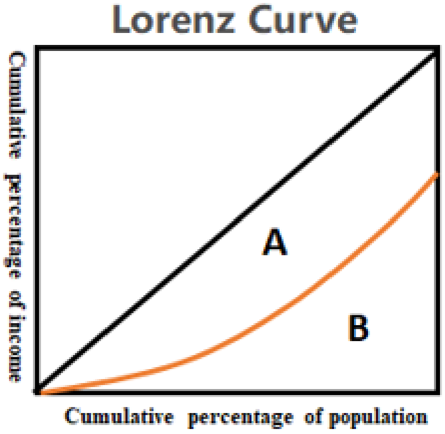

Figure 2. Lorentz Curve Representation for Area Calculation and GINI Index Estimation

Second iteration: Considering $L(x) = 1.875x^2 – 0.875x$. Utilizing points (x=0.8, L(x)=0.500) and (x=1.0, L(x)=1.000) generates the function. Subsequent incorporation of x=0.2, 0.4 and 0.6 produces L(x)= -0.1, -0.05 and 0.15. Upon graphing, a similar outcome emerges – the calculated area A of the Lorentz curve surpasses the actual value, resulting in an evaluated GINI index (See Figure 3).

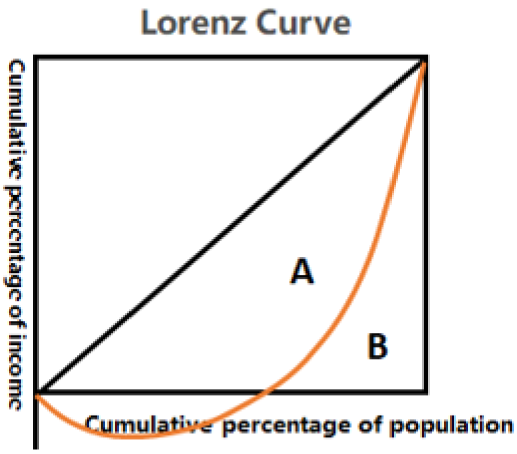

Figure 3. Revised Lorentz Curve Analysis for GINI Index Adjustment

These findings underscore the significance of accurately selecting and fitting a curve to the Lorentz function, as discrepancies can significantly impact the calculation of the Gini Index. Further refinement of curve-fitting methods is crucial for achieving more precise assessments of income distribution.

- Curve-fitting Function Analysis: $L(x) = Ax^n$

In this analysis, we consider the curve-fitting function $L(x) = Ax^n$.

First assumption: Assuming $L(x) = 0.63563x^{1.81943}$ based on data points (x=0.2, L(x)=0.034)and (x=0.4, L(x) = 0.12). Extending the function to x=0.6, 0.8 and 1.0 yields L(x) = 0.250938, 0.423529 and 0.63563.

Observation: Plotting these points on the graph reveals that the calculated area A of the Lorentz curve exceeds the actual value. Consequently, the Gini Index derived from this curve-fitting function is larger than the actual value.

This outcome underscores the critical importance of precision in curve-fitting methods when calculating the Gini Index. Fine-tuning these methods is essential for obtaining more accurate assessments of income distribution and ensuring the reliability of Gini Index calculations (See Figure 4).

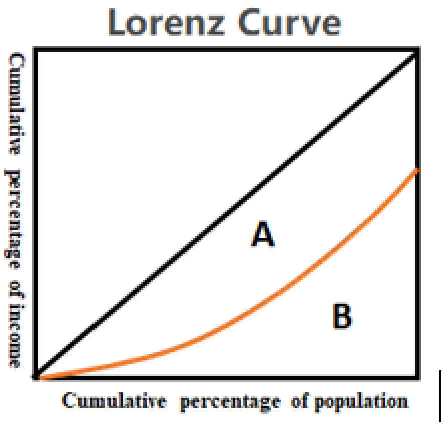

Figure 4. Impact of Curve-Fitting Function L(x) = Axn on GINI Index Estimation

Alternatively, exploring an alternative curve-fitting function, we consider $L(x) = x^{3.10628}$. Assume to apply data points (x=0.8, L(x)=0.500) and (x=1.0, L(x)=1.000) results in the function. Incorporating x=0.2, 0.4, and 0.6 produces L(x) = 0.006742, 0.058061, and 0.204586. We can observe that graphing these points leads to a consistent conclusion: the calculated area A of the Lorentz curve surpasses the actual value.

Thus, the Gini Index derived from this curve-fitting function is larger than the actual value.

This analysis reinforces the critical importance of selecting an appropriate curve-fitting function for accurately calculating the Gini Index. Consistent findings across different functions emphasize the need for refinement in curve-fitting methods to ensure the reliability of Gini Index assessments in the context of income distribution (See Figure 5).

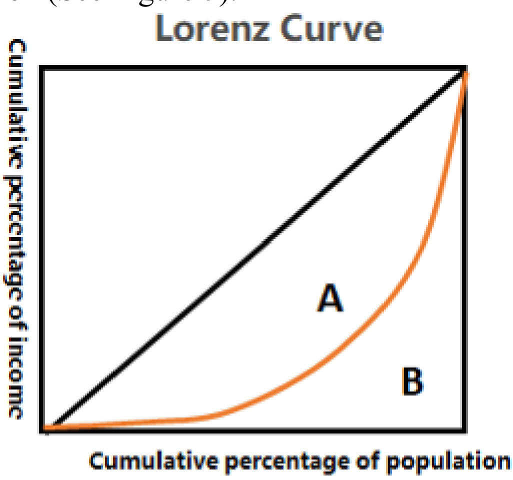

Figure 5. Evaluating the Accuracy of Curve-Fitting Functions in Gini Index Calculation

In summary, using direct curve-fitting functions like $L(x) = Ax^2+Bx$ and $L(x) = Ax^n$ for Gini Index calculation introduces substantial errors. In both cases, the calculated area A of the Lorentz curve consistently surpasses the actual value, resulting in an inflated Gini Index. This highlights the essential need for refined curve-fitting methods to ensure accurate and reliable Gini Index assessments when evaluating income distribution. Further improvements in fitting approaches are crucial for enhancing the precision of Gini Index calculations[6].

- Optimized curve-fitting method with piecewise integral (See Table 2).

Table 2. Piecewise Integral Curve-Fitting Method Results for Gini Index Accuracy

X	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
L(x)	0.000	0.034	0.120	0.267	0.500	1.000

To minimize Gini Index errors, employing the piecewise curve-fitting integral method proves advantageous. In this method, we assume a piecewise quadratic function $L(x) = Ax^2+Bx+C$. This refined approach aims to enhance the accuracy of Gini Index calculations by introducing a more nuanced and adaptable curve-fitting model (See Figure 6).

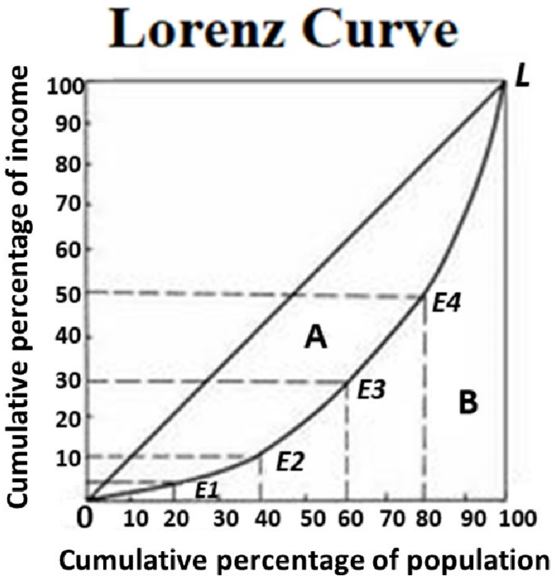

Figure 6. Enhancing Gini Index Accuracy with Piecewise Curve-Fitting Integral Method

First iteration: Using x=0, 0.2, 0.4 and the corresponding L(x)=0.000, 0.034, and 0.12, yields $L(x) = 0.65x^2+0.04x$.

Second iteration: Using x=0.2, 0.4, 0.6 and the corresponding L(x)=0.034, 0.12, and 0.267, we obtain $L(x) = 0.7625x^2-0.0275x+0.009$

Third iteration: Utilizing x=0.4, 0.6, 0.8 and the corresponding L(x)=0.12, 0.267, 0.500, results in $L(x) = 1.075x^2-0.34x+0.084$.

Fourth iteration: With x=0.6, 0.8, 1.0 and the corresponding L(x)=0.267, 0.500, 1.000, we derive $L(x) = 3.3375x^2-3.5075x+1.17$.

These optimized piecewise curve-fitting models provide more accurate representations of the Lorenz curve, contributing to a refined and precise calculation of the Gini Index.

Optimized Results: By employing piecewise integration of the Lorenz curve for the corresponding results 1 and 2, and to minimize errors, it is advisable to calculate the average of the corresponding intervals. The value of area B is determined as the sum of the final results, yielding Size $_{\text{area B}}$ = 0.272503.

This optimization aims to enhance accuracy in the assessment of the Gini Index (See Table 3).

Table 3. Optimization of Lorenz Curve Integration for Gini Index Accuracy

Piecewise	**Result 1**	**Result 2**	**Final result**
0 – 0.2	0.002533		0.002533
0.2 – 0.4	0.014533	0.014383	0.014458
0.4 – 0.6	0.037683	0.037267	0.037475
0.6 - 0.8	0.075267	0.071658	0.073463
0.8 – 1.0	0.144574		0.144574

Finally, we can get the Gini index below.

Gini Index = (0.5 - 0.272503)/0.5 = 0.454994

4.2 Riemann sum method

To enhance precision in Gini Index calculation and scrutinize potential errors, we employ the Riemann sum method in calculus to calculate the area A of the Lorenz curve. For a more meaningful comparison between methods, we maintain consistency by utilizing the same dataset pertaining to income distribution in the United States for the year 2008. This optimized approach aims to provide a comprehensive analysis of the Lorenz curve's impact on Gini Index accuracy (See Table 4).

Table 4. Riemann Sum Analysis of Lorenz Curve

X	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
L(x)	0.000	0.034	0.120	0.267	0.500	1.000

We simplify the analysis by dividing the Lorenz Curve into five equal segments, each corresponding to distinct portions of the cumulative percentage of the population. This streamlined method facilitates a clearer examination of income distribution dynamics, contributing to a more precise calculation of the Gini Index (See Figure 7).

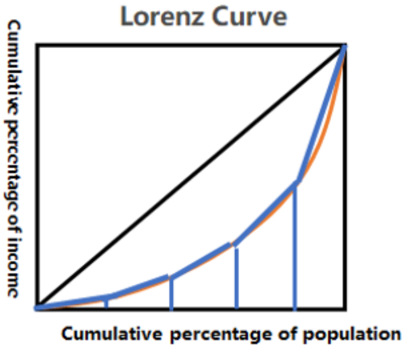

Figure 7. Segmented Lorenz Curve Analysis for Precise Gini Index Calculation

By using the Riemann sum method, we can derive the size of area B.

$$\textbf{Size}_{\textbf{area B}} = \tfrac{1}{2}*[(0+0.034)+(0.034+0.12)+(0.12+0.267)+(0.267+0.5)+(0.5+1)]*0.2 = 0.2842$$

Then we can calculate the Gini index.

$$\textbf{Gini Index} = (1 - 0.2842)/\ (1/2) = 0.4316$$

In conclusion, employing the Riemann sum method in calculus for Gini Index calculation has certain limitations. As the curve is approximated by a straight line when computing area B, the estimated Gini Index tends to be smaller than the actual value. This discrepancy is more pronounced with fewer data points, leading to a larger error.

To mitigate this error, increasing the sample data and subdividing the curve into smaller pieces are recommended. These optimizations enhance the accuracy of Gini Index calculations, providing a more reliable reflection of income distribution dynamics.

4.3 Optimized Gini Index with combined methods

In utilizing the Curve-fitting method for Gini Index calculation, we observe that the area A of the calculated Lorentz curve tends to be larger than the actual value, resulting in an inflated Gini Index. Conversely, when employing the Riemann sum method, the estimated Gini Index tends to be smaller than the actual value, particularly with a limited number of data points, leading to a larger error.

To enhance the precision of our result, we can optimize by calculating the average value of the Gini Index obtained from the two methods. This approach aims to strike a balance between the strengths and limitations of each method, providing a more reliable and accurate assessment of the Gini Index in the context of income distribution.

$$\textbf{Gini Index} = (0.454994 + 0.4316)/2 = 0.443297$$

5. Study on the rate of change and growth rate of Lorenz curve

Let's consider an income distribution function f(x), where x represents income and f represents the number of people corresponding to that income. The integration of function f(x) is denoted as P, representing the total population.

$$\int f(x)dx = P$$

In the second integration, considering the function xf(x) yields I, representing the total income. The cumulative result is obtained by aggregating values from lower to higher income levels.

$$\int xf(x)dx = I$$

Through these integrations, we obtain the cumulative percentage of the population as the horizontal axis and the cumulative percentage of income as the vertical axis, forming a representation of the Lorenz curve.

$$p(x) = \frac{1}{P}\int_0^x f(x)dx$$

$$i(x) = \frac{1}{I}\int_0^x x f(x)dx$$

Consequently, we derive the function i(p) representing the Lorenz curve, where p denotes the cumulative percentage of the population(See Figure 8).

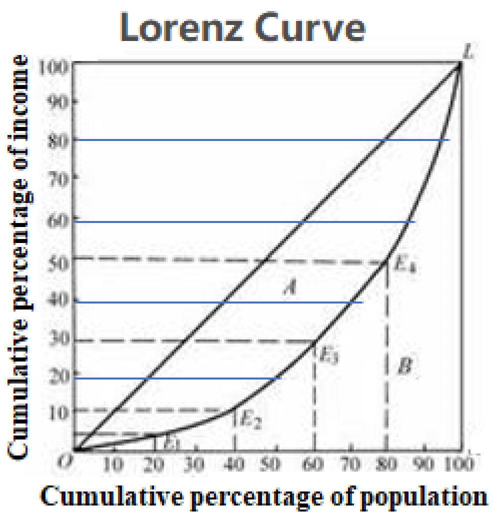

Figure 8. Lorenz Curve Analysis via Function i(p)

Performing the first derivative of the Lorenz curve involves considering $\frac{di}{dp} = \frac{x}{X}$, where X represents the income per capita $X = \frac{I}{P}$. This derivative reveals insights into the relationship between the Lorenz curve and income per capita x. A steeper slope at a specific point on the Lorenz curve indicates a higher income per capita within the corresponding population interval.

Additionally, by extending this analysis to the second derivative,

$$\frac{d^2i}{dp^2} = \frac{1}{X\frac{f(x)}{P}}$$

We further explore the connection between the Lorenz curve and the population f(x) corresponding to income x. A smaller population corresponding to a given income x results in a faster increase in slope.

6. Conclusion

Upon investigating the calculation of the Gini Index and analyzing its change rate and growth rate, certain key conclusions emerge:

- Gini Index and National Income Fairness: The Gini Index serves as an indicator of the fairness in a country's national income. A higher Gini Index suggests a greater disparity in national income.
- Curve-fitting Methods and Error Reduction: Different mathematical models introduce varying errors when estimating the Gini Index. Utilizing the segmented curve-fitting method proves effective in mitigating these errors.
- Optimizing Riemann Sum Method: For the Riemann sum method, error reduction can be achieved by increasing the amount of sample data, thereby minimizing the sample data interval.
- Analyzing Lorentz Curve Changes: Exploring the rate of change of the Lorentz curve and its growth rate enables a deep analysis of the relationship between income per capita and the Lorentz curve's growth rate. This analysis also sheds light on the impact of population changes corresponding to income on the rate of change of the Lorentz curve.

References

[1] Li, Z. X., & Wang, Q. (2019). A Comprehensive Review of the Measurement Methods and Applications of the Gini Coefficient. Journal of Longdong University, 3, 102-105.

[2] Yin, X. H., Li, X., & Yin, C. C. (2021). Equivalence Classification of the Gini Coefficient and Lorenz Curves. Statistics and Decision, 37(24), 5.

[3] Xie, J. (1999). Estimation Methods for Lorenz Curves and Gini Coefficients. Journal of Zhejiang Economy Higher Vocational School, 11(4), 4.

[4] Zhao, Y. B., Wang, J. W., Yin, C., Sun, H., Zhang, L., Fu, Q. (2018). Research on the Equity of Health Resources Allocation in China Based on Lorenz Curves and Gini Coefficients. China Hospital, 022(002), 22-25.

[5] Cheng, Y. H. (2013). The Gini Coefficient of China and Its Decomposition Analysis: Theory, Methods, and Application. China Economics Press.

[6] Zhang, A. G. (2022). A Method for Estimating the Gini Coefficient Based on Grouped Data. Statistics and Decision, 2022(2), 10-15.