Original Research Article

A STUDY OF THE ESTIMATION OF THE GINI COEFFICIENT OF INCOME USING LORENZ CURVE

ABSTRACT

Aims: This paper compares the Boole and Weddle numerical integration methods to estimate the Lorenz curve and Gini Coefficient of income in Ghana.
Study design: Research Paper.
Place and Duration of Study: Ghana, Secondary data for 2013 Ghana Living Standard Survey
Methodology: The Lorenz curve and Gini coefficients of income were estimated using Rasche, Gaffney and Obst function and polynomial function according to numerical integration methods such as Boole and Weddle methods. The Bias and relative error was used to compare the numerical integration methods used.
Results: The results showed that the estimated Lorenz curve and Gini coefficients using Rasche, Gaffney and Obst function and polynomial function according to the Boole and Weddle method of integration resulted in positive and negative biases respectively with the Boole method producing the highest absolute relative error of 1.8082%.
Conclusion: This study showed that both the Boole and Weddle method of numerical integration are not uniformly optimal in estimating the Gini coefficient of income but the Weddle's method is better as compared to Boole method of numerical integration in estimating the Gini coefficient of income.

Keywords: Boole, Weddle, Lognormal, Income, Lorenz Curve, Gini Coefficient

1. INTRODUCTION

Over the years, measuring inequality based on the distribution of income has been a major concern to economist and financial analyst. This has led to an increase in practical and theoretical interest in the distribution of income and how to estimate the Lorenz curve and Gini coefficient of income. Many research has been made to measure variation of income among individuals or household [1, 11]. Using Lorenz curves, the Gini coefficient is defined

as the ratio of the area between the 45^0 line and the Lorenz curve and the area of the whole triangle under the diagonal.

Lorenz curve is used in many fields of life to measure income inequality. Graham [12] considered the inequality of income as a line that separates the rich and the poor. Most of the high income earners reside in the urban areas while majority of the low income earners are seen in the rural areas except for the few who are involved in commercial activities.

Jain [14] inferred that the Lorenz curve is mostly used to measure the economic, political and cultural factors which cause income inequality in the nation. The size of income inequality is mostly measured and interpreted by the Lorenz curve. Mc Donald [18] has shown in his studies that several probability density functions can be used to model income distribution. This is because, the parameters of the probability distribution are easy to estimate and explain. Vital information can be derived directly from estimated parameters when a model which contains parameters is fitted to the income data [2].

The Gamma, Weibull, Exponential, Singh-Maddala, lognormal, Beta, Dagum and Pareto distributions are among the several probability distributions that can be used to model the income distribution. Literature has modelled the income distribution using numerical integration method such as Newton-Cotes, Gauss Quadrature, Romberg Integration and Monte Carlo Integration.

Numerical integration has been applied to several fields such as statistics, actuarial science, engineering finance, etc. [4],[5],[9],[16],[17].

Fellman [10] used the Newton Cotes method such as the Trapezium rule and Simpson's rule to estimate the Gini coefficient of income. He found that the Trapezium rule is the most frequently used numerical integration method for five quintiles. He again found out that for every trapezium rule, the Lorenz curve estimate results in a positive bias and the rule consequently causes negative bias for the Gini coefficient. Newton-Cotes methods such as Trapezium rule, Simpson $1/3$ rule, Simpson $3/8$ rule, Boole's rule and Weddle's rule are special cases of 1st, 2nd, 3rd, 4th and 5th order polynomials used respectively. According to literature, the Weddle's rule is better than the other Newton-Cotes methods but the Boole and Weddle's rule requires a subinterval which is a multiple of 4 and 6 respectively [8].

There are a few or no research in Africa especially in Ghana which estimated the Lorenz curve and Gini coefficient specifically income using functions according to numerical integration. Hence this research seeks to model income distribution of the Ghanaian income and to estimate the Gini coefficient of income using numerical integration. The objective of the study is to compare numerical integration methods such as Boole and Weddle methods used to estimate the Lorenz curve and Gini coefficient.

2. MATERIAL AND METHODS

The distribution of income in the study can be written as a vector $x = (x_1, x_2, x_3, ..., x_n)$ where x_i represent individual income in a population with $i = 1, 2, 3, ..., n$ and n is the overall individuals. The income distribution has mean $\mu(x)$ and density function $f(x)$. Let $\pi(x)$ and $\eta(x)$ represent the cumulative proportion of the individuals that receive income up to x and the cumulative proportion of the total income that is received by individuals in the same population respectively. The Lorenz curve is a graphical representation of the relationship between π and η. Hence;

$$\pi(x) = \int_0^x f(y)dy \tag{1}$$

and

$$\eta(x) = \frac{1}{\mu}\int_0^x yf(y)dy \tag{2}$$

The mean of the distribution of income is;

$$\mu = \int_0^\infty yf(y)dy \tag{3}$$

Chotikapanich and Griffiths [7] stated in their study that a function is fitted to the Lorenz curve to estimate the Lorenz curve. Two functions were selected among the several functional forms proposed for estimating the Lorenz curve for the study. They are a function proposed by Rasche, Gaffney and Obst [19] function and a polynomial function with degree 5. Based on Simulation, a polynomial of degree 5 have proven to be better than polynomials with lesser degrees. Rasche, Gaffney and Obst [19] function which has a nonlinear form is given by;

$$\eta = \left[1 - (1-\pi)^r\right]^{\frac{1}{s}} \tag{4}$$

Where η and π are the parameters and $0 \le r \le 1$ and $0 \le s \le 1$. This Lorenz curve is not symmetric about the $45°$ line perpendicular to the egalitarian line. Generally, the following conditions must be met by the Lorenz curve [15]:

1) If $\pi = 0, \eta = 0$
2) If $\pi = 1, \eta = 1$
3) $0 < \eta < \pi < 1$ (5)
4) The slope of the curve increases monotonically.

The first condition eliminate the chances of an individual receiving zero or negative income. The fourth condition shows that the curve falls below the $45°$ line. The function of the Lorenz curve is always convex and has a constantly increasing positive slope to ensure that the Lorenz curve always fall in the lower triangle of the unit square [19].

Also, Becker and Weispfenning [3] proposed a polynomial which is a mathematical expression involving a sum of powers in one or more variables multiplied by coefficients to estimate the Lorenz curve. A polynomial in one variable (i.e. a univariate polynomial) with constant coefficients is given by;

$$\eta(\pi) = a_m \pi^m + a_{m-1}\pi^{m-1} + \ldots + a_1 \pi + a_0 \tag{6}$$

When $a_m \neq 0$ and $m \ge 2$, the polynomial function is a continuous non-linear function.

The estimate of the area under the Lorenz curve is given by;

$$L(\pi) = \int_0^1 \eta(\pi) d\pi \tag{7}$$

The estimate of the Gini coefficient using the Lorenz curve is given by;

$$\text{Gini} = 1 - 2\int_0^1 \eta(\pi) d\pi \tag{8}$$

The random variable X is modelled using the lognormal distribution with probability density and cumulative distribution function of $f_{(\mu,\sigma^2)}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(Inx-\mu)^2}{2\sigma^2}}, x > 0, \sigma > 0, \mu > 0$ and $F(x) = \phi\left(\frac{Inx - \mu}{\sigma}\right)$ respectively, where Inx is normally distributed. The mean and variance of X is $E(x) = e^{\mu + \sigma^2/2}$ and $V(x) = \left(e^{\sigma^2} - 1\right) e^{2\mu + \sigma^2}$ respectively.

The Gini coefficient of income is computed using the estimated parameters of the lognormal distribution function from the Ghana household data on income would be integrated using the Numerical integration methods such as the Newton- Cotes methods with special emphasis on the Boole rule and Weddle rule. The Newton-Cotes method involves n points in the interval $[a,b]$ with $n-1$ order polynomial which passes through nodes x_i and are equally spaced. Approximating the area under the curve $y = f(x)$ from $x = a$ to $x = b$, using the Boole rule and Weddle rule with $x_0 = a, x_n = b \text{ and } \Delta = \frac{(b-a)}{n}$ is [7];

For Boole rule, we have;

$$A = \frac{2\Delta}{45}\left[7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)\right] \tag{9}$$

For Weddle's rule, we have;

$$A = \frac{3\Delta}{10}\left[f(x_0) + 5f(x_1) + f(x_2) + 6f(x_3) + f(x_4) + 5f(x_5) + f(x_6)\right] \tag{10}$$

The estimates of the Gini coefficients of the various rural and urban areas, regions, male and female heads are compared from the Boole rule and Weddle rule using the relative errors calculated as;

$$\varepsilon_a = \frac{Absolute\ Errors}{Exact} \tag{11}$$

Where $Absolute\ Error = Exact - Approximate$ and the number of significant digits at least correct given as;

$$m = 2 - \log\left(\frac{|\varepsilon_a|}{0.02}\right) \tag{12}$$

The Relative bias is calculated as;

$$R_a = \frac{Absolute\ Bias}{Approximate} \tag{13}$$

Where $Absolute\ Bias = Approximate - Exact$

Approximating the area under the curve, the "exact" numerical integration method was computed using Romberg numerical integration. According to Sauer [20], Romberg integration employs the Composite Trapezoidal method to generate initial approximations and then improve the approximations using the Richardson extrapolation process. The Richardson extrapolation can be executed on any approximation procedure of the form $M - N(h) = K_1 h + K_2 h^2 + ... + K_n h^n$,

where the $K_1, K_2, ..., K_n$ are constants and *N(h)* is an approximation to the unknown value *M.* The truncation error in this formula is dominated by $K_1 h$ when *h* is small, so this formula gives *O(h)* approximations. Richardson's extrapolation employs an averaging method to generate formulas with higher-order truncation error. The Composite Trapezoidal rule for approximating the integral of a function *f* on an interval *[a, b]* using *m* subintervals is

$$\int_a^b f(x)dx = \frac{h}{2}\left[f(a) + f(b) + 2\sum_{j=1}^{m-1} f(x_j) \right] - \frac{(b-a)}{12} h^2 f''(\mu), \tag{14}$$

Where $a < \mu < b, h = \frac{(b-a)}{m}$ and $x_j = a + jh$ for each *j = 0, 1,…, m.*

The Composite Trapezoidal rule approximations are initially acquired with $m_1 = 1, m_2 = 2, m_3 = 4, ..., m_n = 2^{n-1}$, where *n* is a positive integer. The step size h_k corresponding to m_k is $h_k = \frac{(b-a)}{m_k} = \frac{(b-a)}{2^{k-1}}$. With this notation the Trapezoidal rule becomes

$$\int_a^b f(x)dx = \frac{h_k}{2}\left[f(a) + f(b) + 2\left(\sum_{i=1}^{2^{k-1}-1} f(a + ih_k) \right) \right] - \frac{(b-a)}{12} h_k^2 f''(\mu_k), \tag{15}$$

Where μ_k is a number in *(a, b).*

If the Romberg numerical integration, $R_{k,1}$ is introduced to denote the portion of Eq. (15) used for the trapezoidal approximation, then

$$R_{1,1} = \frac{h_1}{2}[f(a) + f(b)] = \frac{(b-a)}{2}[f(a) + f(b)]$$

$$R_{2,1} = \frac{h_2}{2}[f(a) + f(b) + 2f(a + h_2)]$$

$$= \frac{(b-a)}{4}\left[f(a) + f(b) + 2f(a + \frac{(b-a)}{2}) \right]$$

$$= \frac{1}{2}\left[R_{1,1} + h_1 f(a + h_2) \right]$$

$$R_{3,1} = \frac{1}{2}\left\{ R_{2,1} + h_2 [f(a + h_3) + \mathrm{f}(\mathrm{a} + 3h_3)] \right\}$$

Hence, in general,

$$R_{\mathrm{k},1} = \frac{1}{2}\left\{ R_{\mathrm{k}-1,1} + h_{k-1} \sum_{i=1}^{2^{k-2}} f(\mathrm{a} + (2i-1)h_k \right\}, \text{for each } k = 2, 3, ..., n. \tag{16}$$

The Absolute error and bias are necessary to completely characterize the area under the Lorenz curve and Gini coefficient's error since the authors are comparing competing numerical integration techniques.

3. RESULTS AND DISCUSSION

The methodology proposed by this study is fitted to a data on gross income from the Ghana Living Standard Survey (GLSS 6) conducted by the Ghana Statistical Service (GSS). The data comprises of a nationwide sample of 58,788 individuals with a total income of GHC 244,759,213.2. The Ghana Living Standards Survey is a survey conducted throughout the country which assesses the standard of living of the individuals in the population [10]. A Kolmogorov-Smirnov (KS) test was used to test whether the income distribution is lognormal. The null hypothesis of the data modelled on a lognormal distribution, resulted in a p-value of 0.42 greater than 0.05 which is an indication that the income distribution is lognormal.

The total income (x) is modelled using the lognormal distribution with parameters μ and σ^2. The density function of the income which follows the lognormal distribution is given by

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(Inx-\mu)^2}{2\sigma^2}}, x > 0, \sigma > 0, \mu > 0 \qquad (17)$$

The maximum likelihood estimators of the parameters μ and σ^2 of the lognormal distribution based on the income distribution are given as $\hat{\mu} = \frac{\sum_i Inx_i}{n} = 8.5311$ and

$$\hat{\sigma} = \frac{\sum_i (Inx_i - \hat{\mu})^2}{n} = 1.45812$$

The cumulative distribution function of the income is also given by;

$$F(x) = \phi\left(\frac{Inx - 8.5311}{1.45812}\right) \qquad (18)$$

Let $\pi(x)$ and $\eta(x)$ to be the proportion of the units that receive income up to x and the proportion of total income received by the same units whose income are less than or equal to x respectively.

3.1 Estimates of the Gini coefficient using Rasche, Gaffney & Obst [19] method

The estimate of the Gini coefficient of income using the estimate of the Lorenz curve from Rasche, Gaffney & Obst [19] function are shown in Table 1 below;

Table 1: Estimates of Gini Coefficient using Rasche, Gaffney & Obst [19] function

Region/Family Head	Boole Rule	Weddle Rule	"Exact" Integration
Greater Accra	0.47	0.472	0.476
Eastern	0.49	0.494	0.498
Ashanti	0.491	0.4934	0.4996
Volta	0.48	0.4822	0.49
Western	0.53	0.5354	0.54
Brong Ahafo	0.552	0.56	0.568
Central	0.489	0.4916	0.4958
Northern	0.54	0.5408	0.5414
Upper East	0.57	0.578	0.582
Upper West	0.5576	0.5636	0.574
Rural	0.4982	0.4996	0.5008
Urban	0.53572	0.5426	0.55216
Male Head	0.4798	0.4808	0.4818
Female Head	0.55412	0.5614	0.57116
All	0.51696	0.5211	0.52648

Source: Author's research

The computation of the Relative bias and Relative errors using Rasche, Gaffney & Obst [19] function are shown in Table 2. The highest relative error of 2.8571 % was computed using the Boole numerical integration method for the Upper West Region while the lowest relative error of 0.1108% was computed from the Weddle numerical integration method for the Northern Region. Also, the highest relative bias of 1.6673 % was computed using the Boole numerical integration method for the Upper West Region while the lowest relative bias of 0.0600% was computed from the Weddle numerical integration method for the Northern Region.

Table 2: Relative bias and Relative errors of the Gini coefficient using Rasche, Gaffney & Obst [19] function

Region/Family Head	Boole	Rule		Weddle	Rule	
	Relative Bias (%)	Relative Error (%)	m	Relative Bias (%)	Relative Error (%)	m
Greater Accra	0.6036	1.2605	2	0.4016	0.8403	2
Eastern	0.8065	1.6064	2	0.1001	0.8032	2
Ashanti	0.8675	1.7214	2	0.6239	1.241	2
Volta	1.0101	2.0408	2	0.7861	1.5918	2
Western	1.3171	1.8519	2	0.4621	0.8519	2
Brong Ahafo	1.6260	2.8169	2	0.8065	1.4085	2

Central	0.6847	1.3715	2	0.4218	0.8471	2
Northern	0.1402	0.2586	3	0.0600	0.1108	3
Upper East	1.2146	2.0619	2	0.4016	0.6873	2
Upper West	1.6673	2.8571	2	1.0509	1.8118	2
Rural	0.2607	0.5192	3	0.1201	0.2396	3
Urban	1.6715	2.9774	2	0.9652	1.7314	2
Male Head	0.2004	0.4151	3	0.1001	0.2076	3
Female Head	1.7335	2.9834	2	0.9856	1.7088	2
All	0.9612	1.8082	2	0.5409	1.0219	2

**m= the number of significant digits at least correct*

Source: Author's research

3.2 Estimates of the Gini coefficient using Polynomial function with degree 5

The estimate of the Gini coefficient of income for the Boole rule and Weddle rule using the Polynomial function with degree 5 are shown in Table 3 below.

Table 3: Estimates of Gini Coefficient using the Polynomial function with degree 5

Region/Family Head	Boole Rule	Weddle Rule	"Exact" Integration
Greater Accra	0.4694	0.471	0.4754
Eastern	0.4874	0.4944	0.4954
Ashanti	0.488	0.4908	0.4966
Volta	0.4776	0.486	0.4876
Western	0.516	0.5234	0.526
Brong Ahafo	0.491	0.4954	0.567
Central	0.4746	0.487	0.4894
Northern	0.5378	0.5382	0.5392
Upper East	0.569	0.5728	0.581
Upper West	0.5474	0.564	0.5714
Rural	0.4948	0.4966	0.4974
Urban	0.51684	0.528	0.5484
Male Head	0.478	0.4784	0.48
Female Head	0.53364	0.5462	0.5658
All	0.50582	0.5123	0.5229

Source: Author's research

The computation of the Relative bias and Relative errors using the Polynomial function with degree 5 are shown in Table 4. The highest relative error of 2.8571 % was computed using the Boole numerical integration method for the Upper West Region while the lowest relative error of 0.1108% was computed from the Weddle numerical integration method for the Northern Region. Also, the highest relative bias of 2.4590 % was computed using the Boole numerical integration method for the Upper West Region while the lowest relative bias of

0.1001% was computed from the Weddle numerical integration method for the Northern Region.

Table 4: Relative biases and Relative errors of the Gini coefficient using the Polynomial function with degree 5

Region/Family Head	Boole Rule			Weddle Rule		
	Relative Bias (%)	Relative Error (%)	m	Relative Bias (%)	Relative Error (%)	m
Greater Accra	1.6260	1.2621	2	0.4420	0.9255	2
Eastern	0.5025	1.6149	2	0.1001	0.2019	3
Ashanti	0.7354	1.7318	2	0.5834	1.1679	2
Volta	1.1122	2.0509	2	0.1603	0.3281	3
Western	2.3541	1.9011	2	0.2607	0.4943	3
Brong Ahafo	2.6694	13.404	1	0.7212	12.628	1
Central	1.5022	3.0241	2	0.2406	0.4904	3
Northern	0.1402	0.2596	3	0.1001	0.1855	3
Upper East	1.2146	2.0654	2	0.8268	1.4114	2
Upper West	2.4590	4.2002	2	0.7455	1.2951	2
Rural	0.2607	0.5227	3	0.0801	0.1608	3
Urban	3.2588	5.7549	2	2.0825	3.7199	2
Male Head	0.2004	0.4167	3	0.1603	0.3333	3
Female Head	3.3229	5.684	2	1.9992	3.4641	2
All	1.7377	3.2664	2	1.0714	2.0272	2

**m= the number of significant digits at least correct*

Source: Author's research

4. CONCLUSION

The main aim of the research was to compare estimates of Gini coefficient of income for the various regions, family heads and rural and urban area using the Boole and Weddle method of integration. The results from the research showed that the Boole and Weddle methods of numerical integration produced a positive bias for the area under the Lorenz curve using both Rasche, Gaffney & Obst [19] function and the polynomial function with degree 5 while from Table 3, the Gini coefficient yielded a negative bias for both functions. This shows that both the Boole and Weddle method are not uniformly optimal in estimating the Gini coefficient of income using both functions. Also, from the entire income distribution, the Boole method of integration produced the highest relative error and relative bias of 1.8082% and 0.9612% respectively while the Weddle method of numerical integration produced the lowest relative error and relative bias of 1.0219% and 0.5409%respectively using Rasche,

Gaffney & Obst [14] function. Using polynomial function with degree 5, Boole yielded the highest relative error and relative bias of 3.2664% and 1.7377% respectively while Weddle rule yielded the least relative error and relative bias of 2.0272% and 1.0714% respectively. This shows that the Weddle's method of numerical integration is better as compared to Boole method of numerical integration in estimating the Gini coefficient of income.

Future research could apply the Newton- Cotes methods to calculate the Lorenz curve for other non- linear functions such as the polynomial functions or the one proposed by [6].

REFERENCES

1. Aitchison, J., & Brown, J. (1957). The Lognormal Distribution. Cambridge.
2. Alaiz, M.P., & Victoria-Feser, M.-P. (1996). Modelling Income Distribution in Spain: A Robust Parametric Approach: LSE STICERD Research Paper No.20.
3. Becker, T., & Weispfenning, V. (1993). A computational approach to commutative algebra. New York: Springer-Verlag.
4. Burden, R. L., & Fairs, J. D. (2001). *Numerical Analysis, 7th ed.,* Brooks/Cole Thomson, Pacific Grove, CA.
5. Canale, R., & Chapra, S. (2002). *Numerical Methods for Engineers: with Software and Programming Applications*, McGraw-Hill, New York.
6. Chotikabanich D. A (1993) comparison of alternative functional forms for the Lorenz curve, *Economics Letters*, 41, 129-138.
7. Chotikapanich, D., & Griffiths, W. E. (2002). Estimating Lorenz curve using a Dirchlet distribution. Journal of Business and Economic Statistics 20, 290-295.
8. Davis, P. J. and Rabinowitz, P. (1984). Numerical Integration. Academic Press Inc.
9. Dickson, D. C. M., Hardy, M. R. & Waters, H. R. (2013). *Actuarial Mathematics for Life Contingent Risks*, 2nd ed., Cambridge University Press, New York.
10. Fellman, J. (2012). Estimation of Gini coefficients using Lorenz Curves. Journal of Statistical and Econometric Methods; vol.1 no. 2, 31-38.
11. Foster, J. E., & Sen, A. (1997). On economic inequality after a quarter century. Oxford, UK: Clarendon Press.
12. Graham. (1995). Poverty and Health Risks; Issues and Concept.
13. GLSS6. (2014). Poverty profile in Ghana from 2005-2013.*Ghana Statistical Service.*
14. Jain, S. B. (1975). Size Distribution of Income: A computation of Data. Washington: The World Bank.
15. Kakwani, N., & Podder, N. (1973). On the Estimation of Lorenz Curves from Grouped Observations. International Economic Reviews, Vol.14, no.8, 278-292

16. Kaw, A. & Keteltas, M. (2009). Lagrange Interpolation. Retrieved from: http://numericalmethods.eng.usf.edu.

17. Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2004). *Loss Models: From data to Decisions, 2nd ed.*, John Wiley & Sons, Inc. Publication.

18. McDonald, J. B. (1984). Some Generalized Functions for the Size Distribution of Income. Econometrica Vol: 52; No: 3; 647-665.

19. Rasche, R. H., Gaffney, J., Koo, A. Y., & Obst, N. (1980). Functional Forms for Estimating the Lorenz Curve: Econometrica, 48, 1061-1062.

20. Sauer, T. (2012). Numerical Analysis, Second Edition. Pearson Education Inc