

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, 8]. Using Lorenz curves, the Gini coefficient is defined as

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

25

26 Lorenz curve is used in many fields of life to measure income inequality. Graham [9]
27 considered the inequality of income as a line that separates the rich and the poor. Most of
28 the high income earners reside in the urban areas while majority of the low income earners
29 are seen in the rural areas except for the few who are involved in commercial activities.

30

31 Jain [11] inferred that the Lorenz curve is mostly used to measure the economic, political
32 and cultural factors which cause income inequality in the nation. The size of income
33 inequality is mostly measured and interpreted by the Lorenz curve. Mc Donald [13] has
34 shown in his studies that several probability density functions can be used to model income
35 distribution. This is because, the parameters of the probability distribution are easy to
36 estimate and explain. Vital information can be derived directly from estimated parameters
37 when a model which contains parameters is fitted to the income data [2].

38

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

44

45 Fellman [7] used the Newton Cotes method such as the Trapezium rule and Simpson's rule
46 to estimate the Gini coefficient of income. He found that the Trapezium rule is the most
47 frequently used numerical integration method for five quintiles. He again found out that for
48 every trapezium rule, the Lorenz curve estimate results in a positive bias and the rule
49 consequently causes negative bias for the Gini coefficient. Newton-Cotes methods such as
50 Trapezium rule, Simpson $\frac{1}{3}$ rule, Simpson $\frac{3}{8}$ rule, Boole's rule and Weddle's rule are
51 special cases of 1^{st} , 2^{nd} , 3^{rd} , 4^{th} and 5^{th} order polynomials used respectively. According to
52 literature, the Weddle's rule is better than the other Newton-Cotes methods but the Boole
53 and Weddle's rule requires a subinterval which is a multiple of 4 and 6 respectively [6].

54

55 There are a few or no research in Africa especially in Ghana which estimated the Lorenz
56 curve and Gini coefficient specifically income using functions according to numerical
57 integration. Hence this research seeks to model income distribution of the Ghanaian income

58 and to estimate the Gini coefficient of income using numerical integration. The objective of
59 the study is to compare numerical integration methods such as Boole and Weddle methods
60 used to estimate the Lorenz curve and Gini coefficient.

61

62 2. MATERIAL AND METHODS

63

64 The distribution of income in the study can be written as a vector
65 $x = (x_1, x_2, x_3, \dots, x_n)$ where x_i represent individual income in a population with
66 $i = 1, 2, 3, \dots, n$ and n is the overall individuals. The income distribution has mean $\mu(x)$ and
67 density function $f(x)$. Let $\pi(x)$ and $\eta(x)$ represent the cumulative proportion of the
68 individuals that receive income up to x and the cumulative proportion of the total income
69 that is received by individuals in the same population respectively. The Lorenz curve is a
70 graphical representation of the relationship between π and η . Hence;

$$71 \quad \pi(x) = \int_0^x f(y)dy \quad (1)$$

72 and

$$73 \quad \eta(x) = \frac{1}{\mu} \int_0^x yf(y)dy \quad (2)$$

74 The mean of the distribution of income is;

$$75 \quad \mu = \int_0^{\infty} yf(y)dy \quad (3)$$

76 Chotikapanich and Griffiths [5] stated in their study that a function is fitted to the Lorenz
77 curve to estimate the Lorenz curve. Two functions were selected among the several
78 functional forms proposed for estimating the Lorenz curve for the study. They are a function
79 proposed by Rasche, Gaffney and Obst [14] function and a polynomial function with degree
80 5. Rasche, Gaffney and Obst [14] function which has a nonlinear form is given by;

$$81 \quad \eta = \left[1 - (1 - \pi)^r \right]^{\frac{1}{s}} \quad (4)$$

82 Where η and π are the parameters and $0 \leq r \leq 1$ and $0 \leq s \leq 1$. This Lorenz curve is not
83 symmetric about the 45° line perpendicular to the egalitarian line. Generally, the following
84 conditions must be met by the Lorenz curve [12]:

85 1) If $\pi = 0, \eta = 0$

86 2) If $\pi = 1, \eta = 1$

87 3) If $\eta < \pi$ (5)

88 4) The slope of the curve increases monotonically.

89 The first condition eliminate the chances of an individual receiving zero or negative income.

90 The fourth condition shows that the curve falls below the 45° line. The function of the
91 Lorenz curve is always convex and has a constant slope to ensure that the Lorenz curve
92 always fall in the lower triangle of the unit square [14].

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

$$97 \quad \eta(\pi) = a_m \pi^m + a_{m-1} \pi^{m-1} + \dots + a_1 \pi + a_0 \quad (6)$$

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

99 The estimate of the Lorenz curve is given by;

$$100 \quad L(\pi) = \int_0^1 \eta(\pi) d\pi \quad (7)$$

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

$$102 \quad \text{Gini} = 1 - 2 \int_0^1 \eta(\pi) d\pi \quad (8)$$

103 The random variable X is modelled using the lognormal distribution with probability density

104 and cumulative distribution function of $f_{(\mu, \sigma^2)}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$, $x > 0, \sigma > 0, \mu > 0$

105 and $F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$ respectively, where $\ln x$ is normally distributed. The mean and

106 variance of X is $E(x) = e^{\mu + \sigma^2/2}$ and $V(x) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$ respectively.

107 The Gini coefficient of income is computed using the estimated parameters of the lognormal
108 distribution function from the Ghana household data on income would be integrated using
109 the Numerical integration methods such as the Newton- Cotes methods with special
110 emphasis on the Boole rule and Weddle rule. The Newton-Cotes method involves n points
111 in the interval $[a, b]$ with $n-1$ order polynomial which passes through nodes x_i and are

112 equally spaced. Approximating the area under the curve $y = f(x)$ from $x = a$ to $x = b$,
 113 using the Boole rule and Weddle rule with $x_0 = a, x_n = b$ and $\Delta = \frac{(b-a)}{n}$ is [5];

114 For Boole rule, we have;

$$115 \quad A = \frac{2\Delta}{45} [7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)] \quad (9)$$

116 For Weddle's rule, we have;

$$117 \quad A = \frac{3\Delta}{10} [f(x_0) + 5f(x_1) + f(x_2) + 6f(x_3) + f(x_4) + 5f(x_5) + f(x_6)] \quad (10)$$

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

$$121 \quad \mathcal{E}_a = \frac{\text{Absolute Errors}}{\text{Exact}} \quad (11)$$

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

$$124 \quad m = 2 - \log \left(\frac{|\mathcal{E}_a|}{0.02} \right) \quad (12)$$

125 The bias is calculated as;

$$126 \quad \text{Bias} = \text{Approximate} - \text{Exact}$$

127 The "exact" numerical integration method was computed using Romberg numerical
 128 integration. A general expression for Romberg integration can be written as [15];

$$129 \quad I_{2n}^{(k)} = \frac{4^k I_{2n}^{(k-1)} - I_n^{(k-1)}}{4^{k-1} - 1}, k \geq 2 \quad (13)$$

130 The index k represent the order of extrapolation and $I_n^{(1)}$ represent the values obtained from
 131 the regular trapezoidal rule with n intervals. Using $k = 2$ which represents values obtained
 132 using the true estimate as $o(h^2)$.

$$133 \quad \text{Hence } I_{\text{true}} \cong I_{\text{true,est}} = I_{2n} + \frac{I_{2n} - I_n}{3}$$

134

135 3. RESULTS AND DISCUSSION

136

137 The methodology proposed by this study is fitted to a data on gross income from the Ghana
138 Living Standard Survey (GLSS 6) conducted by the Ghana Statistical Service (GSS). The
139 data comprises of a nationwide sample of 58,788 individuals with a total income of GHC
140 244,759,213.2. The Ghana Living Standards Survey is a survey conducted throughout the
141 country which assesses the standard of living of the individuals in the population [10]. A
142 Kolmogorov-Smirnov (KS) test was used to test the normality of the income distribution. The
143 null hypothesis of the data modelled on a lognormal distribution, resulted in a p-value of 0.42
144 greater than 0.05 which is an indication that the income distribution is not normal.

145

146 The total income (x) is modelled using the lognormal distribution with parameters μ and σ^2 .

147 The density function of the income which follows the lognormal distribution is given by

$$148 \quad f_{(\mu, \sigma^2)}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, x > 0, \sigma > 0, \mu > 0 \quad (11)$$

149

150 The maximum likelihood estimators of the parameters μ and σ^2 of the lognormal distribution

151 based on the income distribution are given as $\hat{\mu} = \frac{\sum_i \ln x_i}{n} = 8.5311$ and

$$152 \quad \hat{\sigma}^2 = \frac{\sum_i (\ln x_i - \hat{\mu})^2}{n} = 1.45812$$

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

$$154 \quad F(x) = \phi\left(\frac{\ln x - 8.5311}{1.45812}\right) \quad (12)$$

155

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

159

160 3.1 Estimates of the Lorenz curve, Gini coefficient using Rasche, Gaffney & 161 Obst [13] method

162

163 The estimate of the Lorenz curve for Boole rule and Weddle rule using Rasche, Gaffney and
164 Obst [14] function are shown in Table 1.

165 Table 1: Estimates of Lorenz Curve using Rasche, Gaffney & Obst [14] function

Region/Family Head	Boole Rule	Weddle Rule	"Exact " Integration
Greater Accra	0.265	0.264	0.262
Eastern	0.255	0.253	0.251
Ashanti	0.2545	0.2533	0.2502
Volta	0.26	0.2589	0.255
Western	0.235	0.2323	0.23
Brong Ahafo	0.224	0.22	0.216
Central	0.2555	0.2542	0.2521
Northern	0.23	0.2296	0.2293
Upper East	0.215	0.211	0.209
Upper West	0.2212	0.2182	0.213
Rural	0.2509	0.2502	0.2496
Urban	0.23214	0.2287	0.22392
Male Head	0.2601	0.2596	0.2591
Female Head	0.22294	0.2193	0.21442
All	0.24152	0.23945	0.23676

166 Source: Author's research

167

168 The estimate of the Gini coefficient of income using the estimate of the Lorenz curve from
 169 Rasche, Gaffney & Obst [14] function are shown in Table 2 below;

170

171 Table 2: Estimates of Gini Coefficient using Rasche, Gaffney & Obst [14] 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

172 Source: Author's research

173

174 The computation of the Biases and Relative errors using Rasche, Gaffney & Obst [14]
 175 function are shown in Table 3. The highest relative error of 2.8571 % was computed using
 176 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.

Table 3: Bias and Relative errors of the Gini coefficient using Rasche, Gaffney & Obst [14] function

Region/Family Head	Boole	Rule		Weddle	Rule	
	Bias	Relative Error (%)	m	Bias	Relative Error (%)	m
Greater Accra	-0.006	1.2605	2	-0.004	0.8403	2
Eastern	-0.008	1.6064	2	-0.004	0.8032	2
Ashanti	-0.0086	1.7214	2	-0.0062	1.241	2
Volta	-0.01	2.0408	2	-0.0078	1.5918	2
Western	-0.01	1.8519	2	-0.0046	0.8519	2
Brong Ahafo	-0.016	2.8169	2	-0.008	1.4085	2
Central	-0.0068	1.3715	2	-0.0042	0.8471	2
Northern	-0.0014	0.2586	3	-0.0006	0.1108	3
Upper East	-0.012	2.0619	2	-0.004	0.6873	2
Upper West	-0.0164	2.8571	2	-0.0104	1.8118	2
Rural	-0.0026	0.5192	3	-0.0012	0.2396	3
Urban	-0.01644	2.9774	2	-0.00956	1.7314	2
Male Head	-0.002	0.4151	3	-0.001	0.2076	3
Female Head	-0.01704	2.9834	2	-0.00976	1.7088	2
All	-0.00952	1.8082	2	-0.00538	1.0219	2

*m= the number of significant digits at least correct

Source: Author's research

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

The estimate of the Lorenz curve for Boole rule and Weddle rule using the Polynomial function with degree 5 are shown in Table 4.

Table 4: Estimates of Lorenz Curve using Polynomial function with degree 5

Region/Family Head	Boole Rule	Weddle Rule	"Exact " Integration
Greater Accra	0.2653	0.2645	0.2623
Eastern	0.2563	0.2528	0.2523
Ashanti	0.256	0.2546	0.2517
Volta	0.2612	0.257	0.2562
Western	0.242	0.2383	0.237
Brong Ahafo	0.2545	0.2523	0.2165
Central	0.2627	0.2565	0.2553
Northern	0.2311	0.2309	0.2304
Upper East	0.2155	0.2136	0.2095

Upper West	0.2263	0.2180	0.2143
Rural	0.2526	0.2517	0.2513
Urban	0.24158	0.2360	0.2258
Male Head	0.2610	0.2608	0.26
Female Head	0.23318	0.2269	0.2171
All	0.24709	0.24385	0.23855

Source: Author's research

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 5 below.

Table 5: 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 Biases and Relative errors using the Polynomial function with degree 5 are shown in Table 6. 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.

Table 6: Bias and Relative errors of the Gini coefficient using the Polynomial function with degree 5

Region/Family Head	Boole	Rule		Weddle	Rule	
	Bias	Relative Error (%)	m	Bias	Relative Error (%)	m
Greater Accra	-0.006	1.2621	2	-0.0044	0.9255	2
Eastern	-0.008	1.6149	2	-0.001	0.2019	3
Ashanti	-0.0086	1.7318	2	-0.0058	1.1679	2
Volta	-0.01	2.0509	2	-0.0016	0.3281	3
Western	-0.01	1.9011	2	-0.0026	0.4943	3
Brong Ahafo	-0.076	13.404	1	-0.00716	12.628	1
Central	-0.0148	3.0241	2	-0.0024	0.4904	3
Northern	-0.0014	0.2596	3	-0.001	0.1855	3
Upper East	-0.012	2.0654	2	-0.0082	1.4114	2
Upper West	-0.024	4.2002	2	-0.0074	1.2951	2
Rural	-0.0026	0.5227	3	-0.0008	0.1608	3
Urban	-0.03156	5.7549	2	-0.0204	3.7199	2
Male Head	-0.002	0.4167	3	-0.0016	0.3333	3
Female Head	-0.03216	5.684	2	-0.0196	3.4641	2
All	-0.01708	3.2664	2	-0.0106	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 [14] 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 of 1.8082% while the Weddle method of numerical integration produced the lowest relative error of 1.0219% using Rasche, Gaffney & Obst [14] function. Using polynomial function with degree 5, Boole yielded the highest relative error of 3.2664% while Weddle rule yielded the least relative error of 2.0272%. 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.

235 Future research could apply the Newton- Cotes methods to calculate the Lorenz curve for
236 other non- linear functions such as the polynomial functions or the one proposed by [4].

237
238

239 REFERENCES

240

- 241 1. Aitchison, J., & Brown, J. (1957). The Lognormal Distribution. Cambridge.
- 242 2. Alaiz, M.P., & Victoria-Feser, M.-P. (1996). Modelling Income Distribution in Spain: A
243 Robust Parametric Approach: LSE STICERD Research Paper No.20.
- 244 3. Becker, T., & Weispfenning, V. (1993). A computational approach to commutative
245 algebra. New York: Springer-Verlag.
- 246 4. Chotikabanich D. A comparison of alternative functional forms for the Lorenz curve,
247 *Economics Letters*, 41, (1993), 129-138.
- 248 5. Chotikapanich, D., & Griffiths, W. E. (2002). Estimating Lorenz curve using a
249 Dirchlet distribution. *Journal of Business and Economic Statistics* 20, 290-295.
- 250 6. Davis, P. J. and Rabinowitz, P. (1984). Numerical Integration. Academic Press Inc.
- 251 7. Fellman, J. (2012). Estimation of Gini coefficients using Lorenz Curves. *Journal of*
252 *Statistical and Econometric Methods*; vol.1 no. 2, 31-38.
- 253 8. Foster, J. E., & Sen, A. (1997). On economic inequality after a quarter century.
254 Oxford, UK: Clarendon Press.
- 255 9. Graham. (1995). Poverty and Health Risks; Issues and Concept.
- 256 10. GLSS6. (2014). Poverty profile in Ghana from 2005-2013.*Ghana Statistical Service*.
- 257 11. Jain, S. B. (1975). Size Distribution of Income: A computation of Data. Washington:
258 The World Bank.
- 259 12. Kakwani, N., & Podder, N. (1973). On the Estimation of Lorenz Curves from
260 Grouped Observations. *International Economic Reviews*, Vol.14, no.8, 278-292
- 261 13. McDonald, J. B. (1984). Some Generalized Functions for the Size Distribution of
262 Income. *Econometrica* Vol: 52; No: 3; 647-665.
- 263 14. Rasche, R. H., Gaffney, J., Koo, A. Y., & Obst, N. (1980). Functional Forms for
264 Estimating the Lorenz Curve: *Econometrica*, 48, 1061-1062.
- 265 15. Sauer, T. (2012). Numerical Analysis, Second Edition. Pearson Education Inc
- 266
- 267
- 268
- 269