University of Ghana http://ugspace.ug.edu.gh

UNIVERSITY OF GHANA

APPLICATION OF NUMERICAL INTEGRATION TO STOCHASTIC ESTIMATION OF THE GINI COEFFICIENT

BY

DARKWAH, KWASI ADJEPONG

(10443730)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE MPHIL STATISTICS DEGREE

JULY, 2015

University of Ghana http://ugspace.ug.edu.gh

University of Ghana http://ugspace.ug.edu.gh

University of Ghana http://ugspace.ug.edu.gh

DECLARATION

Candidate's Declaration

This is to certify that, this thesis is the result of my own research work and that no part of it has been presented for another degree in this University or elsewhere.

SIGNATURE: . . . . . DATE. . . . . .

KWASI ADJEPONG DARKWAH

(10443730)

Supervisors' Declaration

We hereby certify that this thesis was prepared from the candidate's own work and supervised in accordance with guidelines on supervision of thesis laid down by the University of Ghana.

SIGNATURE: . . . . . DATE.....

DR EZEKIEL N.N. NORTEY (Principal Supervisor)

SIGNATURE: . . . . . DATE. . . . . .

DR ANANI LOTSI (Co-Supervisor)

University of Ghana http://ugspace.ug.edu.gh

DEDICATION

This thesis is dedicated to my beloved parents Mr. and Mrs. Darkwah, and the rest of my family for the prayers and support throughout my educational endeavour.

University of Ghana http://ugspace.ug.edu.gh

ACKNOWLEDGEMENT

I thank the Almighty God who has given me the care, knowledge and the opportunity to pursue education up to this level. There are many people without whom this work could not have been undertaken. I render my heart-felt thanks to my supervisors; Dr. E.N.N. Nortey and Dr. K. Anani Lotsi for their countless guidance, advice and constructive criticisms throughout this work. I would like to specially thank the Head of Statistics Department, Dr. Felix O. Mettle for his contribution to my work and all the lectures of Statistics Department of University of Ghana, especially Dr. Samuel Iddi, Enoch Quaye, Louis Aseidu and Mr. Abeiku Asare-Kumi for their services and pieces of advice throughout my two years study in this University. To my parents, Mr. and Mrs. Darkwah and siblings, I say thank you all for your support, encouragement, advice and patience throughout my studies and may the good Lord continue to bless you all. I am also grateful to the Carnegie Corporation of New York for its Next Generation of Academics in Africa Project which provided immense support during the course of the study. Finally, to my roommates, Dennis Owusu Adjei and Kwabena Asare and all my friends especially 2015 batch of MPhil students of Statistics Department of University of Ghana, and to all of you including those not mentioned here, I ask for Gods guidance and mercies. Thank you and God bless you.

University of Ghana http://ugspace.ug.edu.gh

ABSTRACT

Over the years, measuring inequality based on the distribution of income has been a major concern to economist. Inequality has had a broader concept than poverty in that it is defined over the entire population not just for the portion of the population below a certain poverty line. The Gini coefficient satisfy many desirable properties of a good measure of inequality such as mean independence, population size independence, symmetry, and Pigou-Dalton Transfer sensitivity. The empirical observation (income) distribution exhibit excess kurtosis and heavy tails. This research first described the probability distribution of income. The study presented a proposed numerical integration method to stochastic estimation of the Gini coefficient. The Proposed Numerical Integration Method showed a better estimate of functions as compared to the Newton's cotes methods such as the Trapezium rule, Simpson's 1/3 rule, Simpson's 3/8 rule, Boole's rule and Weddle's rule. Diagnostic tests such as Q-Q plots and Kolmogorov-Smirnov test were graphically and quantitatively used to assess the fitness to the income data respectively. The study therefore concludes that the proposed method is superior to the Newton-Cotes methods of integration. Also, the Gini coefficient estimate using the proposed numerical integration method with k=3 was 0.48 which shows that there is disparity in income in Ghana and recommend to statisticians or mathematicians to use the proposed numerical integration method when computing functions that can't be easily integrated.

University of Ghana http://ugspace.ug.edu.gh

Contents

1 INTRODUCTION **1**

1.1 Background of the Study 1

1.2 Problem Statement 4

1.3 Objectives of the Study 5

1.4 Scope of the Study 5

1.5 Significance of the Study 6

1.6 Limitations 7

1.7 Organization of the Study 7

2 REVIEW OF SOME RELATED LITERATURE **8**

3 METHODOLOGY **19**

3.1 Introduction 19

3.2 Standard measures of poverty 19

3.2.1 Headcount ratio 20

3.2.2 Poverty gap 21

3.2.3 Squared Poverty Gap Ratio 22

University of Ghana http://ugspace.ug.edu.gh

3.2.4 Poverty Lines 22

3.3 Measures of Income Inequality 23

3.3.1 Lorenz Curve 23

3.3.2 GINI COEFFICIENT 27

3.4 Quantile-Quantile (Q-Q) Plot 28

3.5 Kolmogorov-Smirnov Goodness-of-Fit Test 29

3.6 Probability Distribution of income 30

3.7 Numerical Integration Approach to Find the Area under the Lorenz Curve 32

3.7.1 Trapezoidal Rule 34

3.7.1.1 Error Analysis for the Trapezoid Rule 35

3.7.2 Simpson's 1/3 Rule 37

3.7.2.1 Error Analysis for the Simpson 1/3 Rule . . . 38

3.7.3 Simpson's 3/8 Rule 39

3.7.3.1 Error Analysis for the Simpson 3/8 Rule . . . 40

3.8 Boole's Rule 41

3.8.1 Error Analysis for Boole's Rule 42

3.9 Weddle's Rule 42

3.9.1 Error Analysis for Weddle's Rule 43

3.10 Proposed Numerical Integration formula 43

3.11 Proposed Composite Numerical Integration Method 50

University of Ghana http://ugspace.ug.edu.gh

4 ANALYSIS AND DISCUSSIONS **60**

4.1 Introduction 60

4.2 Empirical trends in Poverty Incidence, Poverty gap and Squared Poverty gap in Ghana from 1991/92 to 2012/2013 60

4.3 Data 61

5 SUMMARY, CONCLUSION AND RECOMMENDATIONS **70**

5.1 Introduction 70

5.2 Summary of Findings 70

5.3 Conclusion 72

5.4 Recommendations 73

5.5 Further Studies 73

University of Ghana http://ugspace.ug.edu.gh

List of Tables

3.1 Estimates of the Area using numerical integration method with 12 subintervals 58

4.1 Descriptive Statistics 62

4.2 Estimates of the Area under the curve and Gini Coefficient of Income in the 10 administrative regions of Ghana 68

University of Ghana http://ugspace.ug.edu.gh

List of Figures

3.1 The distribution and disparity of income 25

3.2 Monotonically Increasing function of $f(x)$ when $k = 1$ 46

3.3 Monotonically decreasing function of $f(x)$ when $k = 1$ 47

3.4 Monotonically Increasing function of $f(x)$ when $k = 2$ 48

3.5 Monotonically decreasing function of $f(x)$ when $k = 2$ 49

4.1 Proportional Changes in Poverty Incidence, Poverty Gap and Squared Poverty Gap in Ghana from 1991 to 2013 61

4.2 Histogram of Incomes 63

4.3 Lognormal Q-Q plot 64

4.4 The Lorenz Curve of income 65

INTEGRI PROCEDAMUS

University of Ghana http://ugspace.ug.edu.gh

Chapter 1

INTRODUCTION

1.1 Background of the Study

Over the years, measuring inequality based on the distribution of income has been a major concern to economist. Comparing inequality to poverty, the concept of inequality is broad in extent than poverty. This is because the definition of inequality is based on the whole population and not only a section of the population below a definite poverty line. Several measures of inequality must satisfy six basic criteria to become an accepted measure of inequality. They are; statistical testability, mean independence, Pigou-Dalton transfer sensitivity, population size independence, symmetry and decomposability (Glewwe, 1985).

The simple inequality measure of a population separates those who are poor from those who are rich and shows the percentage of income ascribed to every quintile or decile of the population. The quintile which shows those who are poor are mostly 6–10 percent of all income and the highest quintile fall within 35–50 percent. Gini (1912), Theil (1967), Kakwani (1980), Chakravarty (1988), Basman and Slottje (1988) have all devised various ways of measuring

University of Ghana http://ugspace.ug.edu.gh

income inequality. The importance of measuring income on national basis is to ensure an efficient management of the financial and economic policies that will ensure price stability and economic growth.

Graham (1995) 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. According to Edozien (1991), poor health, illiteracy, unemployment and poverty are some characteristics of people who receive low income while literacy, stable jobs, sufficient health care are the characteristics of high income earners.

Jain (1975) 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. The Lorenz curve is a graph that accounts for the cumulative percentage of household income to the cumulative percentage of income received when the households are put in ascending order of their income. The equation or polynomial function of the Lorenz curve can be generated from the probability distribution function of the income (Kakwani & Podder, 1973).

Goldie (1977) also studied the asymptotic framework of the Lorenz curve in what nowadays would be called an empirical-process framework. Gini coefficient is one of the most efficient and commonly used measurement of income inequality in the world. It is computed as twice the region between a Lorenz curve and the egalitarian line. Accurate estimates of the Gini coefficient is computed from primary income data (Fellman, 2012). The Gini coefficient satisfies the Pigou-Dalton Transfer sensitivity, mean independence, symmetry, and population size independence properties of inequality (Glewwe, 1985). Gini coefficient as a measure of inequality is widely used in various contexts

University of Ghana http://ugspace.ug.edu.gh

such as energy, credit availability, income, health care and wealth (Berndt et al. 2003). This indeed confirms Morgan's (1962) statement that the Gini index is the best single measure of inequality.

A number of alternative probability density functions have been considered as models of income distribution (McDonald, 1984). Such models have supplied close estimates to the original distribution, and the parameters of the probability distribution are easy to be estimated and interpreted. Fitting a model that contain parameters to income data can be very useful in extracting important information directly from the parameters which are estimated (Alaiz & Victoria-Feser, 1996). These probability distributions include, among others, the exponential, lognormal, gamma, beta, Singh-Maddala, Dagum, Pareto and Weibull distributions. Literature has investigated the extreme values or whole range of the income distribution (Cowell, 1977).

Several theories that have explicated the distribution of individual income have emanated from two main schools of thought. The stochastic theory of distribution originated from Aitchison and Brown (1957), Mandelbrot (1960), Steindl (1965) and which is the first school of thought. These scientists used stochastic processes to explicate the generation of income showing that in stochastic processes, the distribution is stationary. The socioeconomic school which is the second school of thought explicates the individual income distribution using economic factors such as occupation, distribution of wealth, age, geographical differences, sex and education.

The Gini coefficient is analyzed based on discrete and continuous distributions (Yitzhaki & Schechtman, 2005). Golden (2008) demonstrated how estimates of Gini coefficient can be computed using numerical integration method. Gastwirth (1972), Mehran (1975), Needleman (1978) and Fellman (2012) used the Newton Cotes such as Trapezium rule and Simpson's rule to estimate the Gini

University of Ghana http://ugspace.ug.edu.gh

coefficient of income.

1.2 Problem Statement

Recently, there has been an increasing practical and theoretical interest in the distribution of income and how to estimate the Lorenz curve and Gini coefficient of income. There are a lot of literature on the functions of income distribution that can be easily integrated but very few literature on functions that cannot be integrated analytically. Over the years, numerical integration methods such as Newton-Cotes, Romberg Integration, Gauss Quadrature and Monte Carlo Integration are used for this integration. Newton-Cotes methods such as Trapezium rule, Simpson 1/3 rule, Simpson 3/8 and Boole's rule are special cases of 1st, 2nd, 3rd and 4th order polynomials used respectively.

Several efforts have been made to achieve the exact estimates of the Lorenz curve and Gini coefficient. This has led to numerous criticisms about the computation of the Gini coefficient as a measure of income inequality using probability distribution functions such as the exponential, Pareto distribution etc. The trapezium rule is easier to compute but the area under the Lorenz curve produces a positive bias which results in a negative bias for the Gini coefficient. Simpson´s 1/3 rule which requires an even number of subintervals is better fitted to the Lorenz curve as compared to the Trapezium rule. Lagrange polynomials of second degree whose subintervals are equal can be regarded as the same as Simpson´s 1/3 rule.

Furthermore, there has been a few studies on numerical integration methods used to estimate Lorenz curve and Gini coefficient specifically income, in Africa and Ghana in particular. Also, even though the Gini coefficient has been employed successfully to various data in different countries, mostly in the

University of Ghana http://ugspace.ug.edu.gh

developed areas, there is insufficient empirical evidence of its performance in the developing and the underdeveloped parts of the world.

This study propose a new numerical integration method that is an extension of the trapezium rule and uses it in the estimation of the Lorenz curve and Gini coefficient. That is, this proposed numerical integration method can be used for both even and odd sample sizes. Hence this study seeks to address the problem of modelling income distribution of the Ghanaian income and computing the Gini coefficient of income.

1.3 Objectives of the Study

The main objective of this study is to propose a numerical integration method that provides better estimates as compared to the Newton-Cotes methods of integration, and apply it to the stochastic estimation of the Gini coefficient.

The study specifically seeks to;

1. Propose a numerical integration method.
2. Compare the proposed numerical integration method to the Newton-Cotes methods of integration, i.e. the Trapezium rule, Simpson $1/3$ rule, Simpson $3/8$ rule.
3. Apply the proposed numerical integration method to the stochastic estimation of the Lorenz curve and Gini coefficient.

1.4 Scope of the Study

Research over the years have seen analysts departing from the usual traditional approaches in computing the estimates of the Lorenz curve and Gini coeffi-

University of Ghana http://ugspace.ug.edu.gh

cient to using numerical integration methods. This transition has however not produced a lot of application in the African context. In Ghana, the Gini coefficient has rarely been used to measure income inequality. The proposed numerical integration method is used to estimate the Lorenz curve and Gini coefficient. The study will consider the gross income from the Ghana Living Standard Survey conducted in 2013.

1.5 Significance of the Study

Although there are several research in measuring the income inequality using Lorenz curve and Gini coefficient, there are a few or no studies that have used numerical integration methods to compute the area under the Lorenz curve. Findings from this study will aid economists, financial analysts and policy makers in making economic policies that will produce positive results in all the sectors of the economy. The results from this study will also be very useful in assisting policy makers, stakeholders, investors and financial analysts to employ the proposed numerical integration method in computing the income inequality of a country whose income data follows a probability distribution. It is further suggested that statisticians or mathematicians who want to integrate a function that can't be computed analytically could use the proposed numerical integration method. On the whole, it will add knowledge to the academic field, since this is a new numerical integration method that extends the trapezium rule and there is little or no work that has been carried out in Ghana.

University of Ghana http://ugspace.ug.edu.gh

1.6 Limitations

The Proposed Numerical Integration Method is simple and gives accurate estimates in definite integration but requires alot of computation of the ordinates of a higher level of the subintervals.Nevertheless, the help of computers will enable us to develop an algorithm which should be able to compute the numerical integration estimate efficiently and with ease.

1.7 Organization of the Study

The remaining thesis is organized as follows; chapter two is a review of some of the literature related to the work and chapter three presents an in-depth discussion of the methodology employed by looking at the distribution of income, measures of income inequality such as Gini coefficient, various numerical integration methods and goodness of fit tests of lognormal distribution. Chapter four analyzes the data collected and presents a detailed discussion of the observed results and finally chapter five summarizes the findings, concludes the study and submit some recommendations based on the findings of the study.

INTEGRI PROCEDAMUS

University of Ghana http://ugspace.ug.edu.gh

Chapter 2

REVIEW OF SOME RELATED LITERATURE

A number of alternative probability density functions have been considered as models of income distribution (McDonald, 1984). Such models have supplied close estimates to the original distribution, and the parameters of the probability distribution are easy to estimate and interpret. Fitting a model that contains parameters to income data can be very useful in extracting important information directly from the parameters which are estimated (Alaiz & Victoria-Feser, 1996). These probability distributions include, among others, the exponential, lognormal, gamma, beta, Singh-Maddala, Dagum, Pareto and Weibull distributions. Literature has investigated the extreme values or whole range of the income distribution (Cowell, 1977).

Several theories that have explicated the distribution of individual income have emanated from two main schools of thought. The stochastic theory of distribution originated from Aitchison and Brown (1954), Mandelbrot (1961) and Steindl (1965) which is the first school of thought.These scientists used stochastic processes to explicate the generation of income showing that in stochastic

University of Ghana http://ugspace.ug.edu.gh

processes, the distribution is stationary (Dorfman, 1979). The socioeconomic school which is the second school of thought explicates the individual income distribution using demographic, political and economic factors such as occupation, distribution of wealth, age, geographical differences, sex and education (Levy, 1987).

The Lognormal and Pareto distributions are frequently used as models of income distributions. Pareto (1897) proposed that the universal power law is always substantial to model income distribution of countries. Shirras (1935) disputed this conjecture and inferred; "There is indeed no Pareto Law. It is time it should be entirely discarded in studies on distribution". A "weak Pareto law" suggested by Mandelbrot (1960) can only be applied asymptotically to incomes which are high. This shows that Pareto's proposal is futile to represent a greater portion of the population. The data is fitted better for high incomes using the Pareto function as compared to low incomes but the lognormal distribution is considered when the whole income distribution of a country is used (Cramer,1978).

The findings of Gibrat (1931) in a study revealed that the practical description of lognormal density function could be ascribed based on three requirements; (i) individual income received at a given time is based on previous income and a random element (ii) the random element is a function of a large number of small factors, and (iii) every factor in the population is independent of another. According to the Central Limit Theorem, the log of income based on those requirements is made up of aggregate of several independent random variables which is approximately normal. Gibrat (1931) also established that income is influenced by a multiplicative random effect, which concludes as a lognormal distribution. Further research was conducted on Gibrat's research work. Kalecki (1945) found that "the variance of log income was relatively

University of Ghana http://ugspace.ug.edu.gh

constant". This influenced his decision that when income is low, the probability of shocks which are negative is small. Steyn (1959) ascertained that the income distribution of people who reside in rural and urban areas could be modelled by a lognormal distribution.

Sargan (1957) generated accurate models that produce lognormal distributions of earnings. Balintfy and Goodman (1973) confirmed that lognormal distribution can be used to explicate income distribution since income can be modelled as a random variable. Souma (2000) identified that the lognormal distribution gives a good fit to income distribution at lower income levels using a large data set of resident income. Sala-i-Martin (2006) used the kernel density estimator which is a mixture of 5 lognormal. The conclusion can be interpreted using different stochastic processes with estimates of variances for every twentieth percentile. Cheng (2005) compared the distribution of socio-economic factors to the lognormal distribution and found that the lognormal distribution is universal of the socio-economic factors and made argument that the standard deviation was used to generate the Gini coefficient under fairly general conditions of the lognormal distribution.

In 2006, the World Bank showed empirically that the distribution of income follows the lognormal distribution using 40 years residents' income of both developed and developing countries (Lopez & Servén, 2006). Also, it is empirically proven by López and Servén (2006) in a study of 800 country-year observations that per capita income follows a lognormal distribution. The lognormal distribution was utilized by Holzmann, Vollmer and Weisbrod (2003) to fit income distribution. It was used to examine the inequality of global income and poverty. Pinkovskiv (2008) also analyzed 37 dataset and realized that the lognormal mixtures employed, perform better than the generalized beta functions.

University of Ghana http://ugspace.ug.edu.gh

Anand (1983) suggested that for any index of income inequality, the following characteristics should be satisfied; population-size independence, mean or scale independence and Pigou-Dalton condition. Some of these measures of income inequality are; interquartile range, Variance, Standard deviation, Theil entropy index and Gini coefficient. Wilmoth and Horiuchi (1999) , Anand and Nathikesan (2000) and Anand et al. (2001) considered some of these measures of inequality which have been used to analyze the distribution of income based on the characteristics for any index of income inequality. The interquartile range satisfies all the conditions except the Pigou-Dalton condition (Wilmoth & Horiuchi, 1999). The population-size independence and Pigou-Dalton condition is satisfied by variance and the standard deviation of income but not mean independence. This means that the mean value can change with the measurement of inequality remaining constant even if the distribution does not changed (Anand, 1983). The Theil entropy index is dependent on the entropy in the theory of information (Anand, 1983 ; Theil, 1967). This satisfies all the characteristics of inequality. Comparing it with Gini, Inter-quartile range and Standard deviation, Theil entropy is not easy to understand and interpret. Gini coefficient satisfies the three characteristics of inequality and, therefore demonstrates the importance of the Lorenz curve among distributions of income (Anand, 1983; Goodwin & Vaupel, 1985) and its as well easy to interpret.

The first inequality index was developed by the Italian statistician Gini (1912) and is called the Gini index or Gini coefficient. It is the principal measure of inequality commonly used in Economics and has recently been used in many fields of studies. The Gini index wasn't originally developed to analyze the inequalities of income but to provide a statistic for showing the variations of a random variable. Later, Dalton (1920) used the index in his studies when

University of Ghana http://ugspace.ug.edu.gh

analyzing the distribution of income and the index was named Gini coefficient. Gini (1912) defined the index as;

$$\text{Gini} = \frac{1}{2m^2\bar{z}} \sum_{i=1}^{m} \sum_{i=1}^{m} |z_i - z_j| \tag{2.1}$$

where $z_i$is the proortion of income earned by $i = 1, 2, ..., m$ individuals and $\overline{z}$ is the mean income earned.

Furthermore, the variation of either the income distribution, wealth, consumption or any other type of probability distribution can be measured using the Gini index which is seen as a function of the mean difference by statisticians (Xu, 2004). It is mostly used in the field of economics because of its intuitive geometric interpretation. That is, it is twice the area between the line of perfect equality and the Lorenz curve and it is a necessary element of the Index of poverty intensity developed by Sen (Xu and Osberg, 2002).

The Lorenz curve that graphically shows the distribution of income in a nation is defined by Sen (1973) as a curve that displays the cumulative proportion of the population on the horizontal axis and the cumulative proportion of income enjoyed by the population on the vertical axis. Kendall and Stuart (1958) also introduced the Lorenz curve by using a positive random variable to represent income. Lorenz curve has shown it's importance in other fields of science including measuring differences in the population of plants and estimating the inequality in individual health even though it was mostly used to measure disparities in the distribution of income (Damgaard & Weiner, 2008).

Theoretically, the discrete and continuous distributions are the two main ways for estimating the Gini index. Dorfman(1979) stated that both approaches can be merged and Yitzhaki and Schechtman (2005) confirmed that the continuous distribution is easier to formulate as compared to the discrete random

University of Ghana http://ugspace.ug.edu.gh

variable. Needleman (1978) investigated various ways to produce more accurate estimates for Gini coefficient. Gastwirth (1972) presented an exact interval estimate of the Gini coefficient and compared the Gini with Tepping's findings to assess the exactness of the estimates. Mehran (1978) suggested an alternative estimation method for interval estimation of Gini coefficient which showed that, the grouping limits of Tepping's data are not equidistant, yields a negative bias of the trapezium rule and cannot be applied to the Simpson's rule. Needleman (1978) started his study by investigating the Gini coefficient using the Trapezium rule $G_L$and then incorporated an upgraded upper estimate G_U. The estimate of his method follows the "two-thirds rule" that is;

$$G = \frac{G_L}{3} + \frac{2G_U}{3} \tag{2.2}$$

Lorenzen (1980) presented information on household income distribution and showed that the Gini coefficient gave negative biases using the trapezium rule. Golden (2008) demonstrated how estimates of Gini coefficient can be computed using numerical integration method. McDonald and Ransom (1981) applied Monte Carlo methods to Gamma density to estimate the lower and upper limits of the Gini coefficient. Fellman (2012) 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.

Gastwirth (1972) also revealed that the Lorenz curve can better be estimated by Simpson´s rule as compared to the trapezium rule since it gives accurate results for quadratic and cubic curves but the Simpson's rule demands an even

University of Ghana http://ugspace.ug.edu.gh

subinterval. Berrut and Trefethen (2004) also showed in their studies that indeed the subintervals of Simpson's rule should be even and concluded that the Lagrange polynomials of second order can be considered as a generalization of Simpson´s rule. Berrut and Trefethen (2004) also confirmed that the Lagrange polynomial method is identical to Simpson's rule and hence the polynomial attained has to be integrated to produce approximated Gini coefficients.

Giorgi and Pallini (1987) investigated various ways to produce more accurate estimates for Gini coefficient by including theoretical studies such as Pareto distribution in the study. They found out that among all the estimates of Gini coefficient, the largest errors were produced by the trapezium rule and that the Simpson's rule is uniformly the same as the Golden rule. Garner (1993) used the Lerman and Yitzhaki covariance method to analyze inequality in the distribution of household consumption expenditures, and to examine relationships between various expenditure budget components and total expenditures using United States data. This method had previously been applied to the study of income inequality by income source in the United States of America (Lerman & Yitzhaki, 1984). It was also used to assess the progressivity of taxation in Israel (Yitzhaki, 1990), and to determine the welfare dominance of excise taxation for Cote d'Ivoire (Yitzhaki and Thrisk, 1990).

Other researchers used the Gini coefficient and concentration curves to produce income or expenditure elasticity (Iyengar, 1960; Kakwani, 1978; Blaylock & Smallwood, 1982; Yitzhaki, 1990; Yitzhaki & Thrisk, 1990). In each of these previous studies, with the exception of Kakwani (1978), elasticity was produced for a selected few commodities and commodity groups. Gartner's (1993) study disaggregated total expenditures into nine exhaustive categories; food, shelter, household fuels and utilities, household operations, apparel and services, transportation, medical care and services, entertainment, and other

University of Ghana http://ugspace.ug.edu.gh

expenditures. Expenditures pertaining to savings such as personal insurance and pensions were excluded. Micro level data from the United States of America Consumer Expenditure Interview Survey were analyzed with expenditures referring to those of the consumer unit (United States Department of Labour, 1978).

On another front, Silber (1989) presented the use of G-matrix, which greatly simplifies not only the computation of the Gini Index but also the decomposition by factor components or population subgroups. The proposed approach also allows one to give a clear interpretation of the interaction term which is obtained when the Gini Index is broken down by population subgroups. According to Sen (1973) and Donaldson and Weymark (1981), the Gini index of inequality G_I could be written as;

$$G_I = \sum_{i=1}^{m} y_i \frac{(m-i)}{m} - \frac{(i-1)}{m} \tag{2.3}$$

Where y_i the proportion of total income earned by the individual whose income has the i^{th} rank in the income distribution, assuming that

$y_1 \geq y_2 \geq ... \geq y_i \geq ... \geq y_m$

Expression (2.3) may be written as (Silber, 1989)

$$G_I = \sum_{j=1}^{m} y_i \left[\sum_{i \geq 1} \left(\frac{1}{m} \right) - \sum_{i \leq 1} \left(\frac{1}{m} \right) \right] \tag{2.4}$$

According to Berrebi and Silber (1987) equation (2.4) can be written as;

$$I_G = (e\prime Gs) \tag{2.5}$$

Where e is a column vector of m elements which are equal to $1/m$ ($e\prime$ and s

University of Ghana http://ugspace.ug.edu.gh

are the row and column vectors of m elements which equals $s_1, s_2, ..., s_n$and G (which could be called the G-matrix) is an $n\chi n$ matrix whose elements g_{ij} are equal to -1 when $j > I$, to $+1$ when$I > j$ and 0 when $I = j$. This approach may be used to estimate an upper bound to the Gini Index when only group observation are available (Silber, 1989). Husted (1991) estimated the Gini coefficient by studying the changes in state-level income inequality. The study uses techniques developed in Kakwani and Podder (1973) and described in Slottje (1989) to estimate state-level Lorenz curves. The parameter estimates were then used to calculate the associated Gini coefficients and their asymptotic standard errors. Following Kakwani and Podder (1973), when family income is a random variable with probability density function, then any point of a Lorenz curve can be described by two line segments,

$$\pi = \frac{1}{\sqrt{2}}\left[F(z) + K(z)\right], \sigma = \frac{1}{\sqrt{2}}\left[F(z) - K(z)\right] \tag{2.6}$$

Where $F(z) = \int_0^z f(z)dz$ is the proportion of families with income less than or equal to z and $K(z) = \frac{1}{\mu}\int_0^z zf(z)dz$ is the proportion of income earned by the families who have income less than or equal to z , where μ represents the mean income of all families in the population. According to Husted (1991), the Lorenz curve for each state income is;

$$\sigma = a\pi\alpha(\sqrt{2} - \pi)^{\beta}, 0 < \pi < \sqrt{2}, a, \alpha, \beta > 0 \tag{2.7}$$

The Gini coefficient for each state derived from the Lorenz curve in equation (2.7) is (Husted, 1991);

$$G = 2a(\sqrt{2})^{1+\alpha+\beta}\beta(1 + \alpha, 1 + \beta) \tag{2.8}$$

University of Ghana http://ugspace.ug.edu.gh

Where $\beta(1+\alpha, 1+\beta)$ is the beta function, with $1+\alpha$ and $1+\beta$ degrees of freedom. Husted (1991) calculated state level Gini coefficients for family income from 1981 to 1987 and reflected the increasing inequality observed in the United States during the 1980s. This measure also indicated significant differences in the growth of family income inequality across states and the uneven incidence of the state level economic restructuring, from high and middle wage manufacturing, construction, and mining sectors to the low wage service and trade sectors (Husted, 1991). Remarkable work has been done by Kratz and Resnick (1995) in the area of the usual graphical tools employed in the preliminary data analysis. They present a detailed analysis and discussion of results obtained from using Quantile-Quantile (Q-Q) plots.

Embrechts, Kluppelberg and Mikosch (1997) also offered a very detailed analysis of Q-Q plots. The fitting of a lognormal distribution to a data can be described graphically using exploratory data analysis. Nolan (2003) argues that the P-P plots allows a better graphical examination of the entire range of a data set but the Q-Q plot cannot graphically examine heavy tailed data better and hence suggested the "variance stabilized" P–P plot developed by Michael (1983). A finite population's variance have lesser values which fall within the extreme than a data set which is heavy tailed. This condenses the Q-Q plot, allowing few values which are extreme to dominate the plot (Nolan, 2003). The univariate quantile-quantile (Q-Q) plot is frequently used to examine whether two independent distributions are the same or different (Gnanadesikan & Wilk, 1968). This study displays the graphical methods discussed in examining the income distribution.

The above discussion has mainly highlighted the income distribution. It reviews a range of income data that have been modelled with the lognormal distribution using different numerical integration methods to calculate the Lorenz

University of Ghana http://ugspace.ug.edu.gh

curve. The chapter also reviews different methods for estimating the Gini coefficient. The next chapter elaborates further on this statistical technique and proposed a numerical integration method that gives a detailed account of finding the area under a curve which is better than other Newton-Cotes methods.

University of Ghana http://ugspace.ug.edu.gh

Chapter 3

METHODOLOGY

3.1 Introduction

This chapter presents a detailed description of the data collected and discussion of the various methods used in the analysis of the data. A comprehensive review of the theoretical background underlying the standard measures of poverty and various numerical integration approaches such as Trapezium rule, Simpson's 1/3 rule, Simpson's 3/8 rule in estimating the Lorenz curve and Gini coefficient applied to a probability distribution. Specifically, this chapter examines the lognormal distribution. Techniques for estimating the parameter of the lognormal distribution as well as Quantile-Quantile (Q-Q) plot used in assessing goodness of fit to the data set is described. The Kolmogorov-Smirnov goodness of fit test applied in this research work is also discussed.

3.2 Standard measures of poverty

There are generally two main processes in assessing poverty; identifying and grouping those who are poor to measure them (Sen, 1984). The first process

University of Ghana http://ugspace.ug.edu.gh

is achieved by using any of the measures of poverty. This is made up of separating those who are poor from those who are not poor. The next process is grouping those who are poor to form a single unit. Many indices measuring those who are poor using the poverty line are available but in practice, the most commonly used measures of poverty are the headcount index, the poverty gap index and the severity of poverty index. Foster, Greer and Thorbecke family of measuring poverty is made up of these three measures of poverty (Foster, Greer & Thorbecke, 1984). This is defined as;

$$w_\beta = \frac{1}{m}\sum_{j=1}^{r}\left(\frac{k - z_i}{k}\right)^\beta \tag{3.1}$$

Where classifying of individuals start from those who are poor $(j = 1)$ to those who are rich $(j = m)$ where m is the size of the population. Given z_j as the income, k as the poverty line, r as the number of economic units showing the burden on government to cater for those who are poor, then when $\beta = 0$, the index decreases to a headcount, when $\beta = 1$, it decreases to the poverty gap and when $\beta = 2$, it decreases to the squared poverty gap.

3.2.1 Headcount ratio

The Headcount ratio, w_0 also referred to as the incidence of poverty is the easiest and most common measure of poverty in the world. It is given by the proportion of individuals in a population who fall below the stipulated poverty line k. That is the proportion of individuals in a population that are ranked as poor. It is very easy to compute and explain. Headcount ratio meets the requirements of focus axiom and is additively decomposable. It is computed by comparing every household income z_j to the poverty line k. Given $j = 1, 2, 3, ..., m$ where m is the number of individual households in the

University of Ghana http://ugspace.ug.edu.gh

sample, then each household is formed using an indicator variable with the value 0 representing income above the poverty line or 1 if income is lesser. That is;

$$I(s,t) = 1, if\ z_j \leq k \tag{3.2}$$

and

$$I(s,t) = 0, if\ z_j \geq k. \tag{3.3}$$

The headcount ratio is the sample mean of the variable (s,t), weighted by the number of individuals in every household m_i. The headcount ratio is given by;

$$w_0 = \frac{r}{m} \tag{3.4}$$

where $r = \sum_{i=1}^{h} I(s,t)m_i$ is the number of individuals who are classified as poor and $m = \sum_{i=1}^{h} m_i$ is the total number of individuals in a population.

The Headcount ratio considers those who are poor but not the intensity of their poverty. Rowntree (1901) have broadly used the head count ratio as a measure of poverty ever since it was developed.

3.2.2 Poverty gap

Poverty gap was defined by Sen (1973) as the collection of shortage of income of all the poor based on the value of the poverty line k, weighted by the rate of those who are poor. Given the poverty line k, z_j being the income of person j and m being the population size, then income gap of every individual is given by;

University of Ghana http://ugspace.ug.edu.gh

$$P_j = k - z_j. \tag{3.5}$$

The poverty gap ratio shows the mean distance between the income of those who are poor and the poverty line. Therefore the poverty gap ratio considers both the rate of occurrence and the intensity of those who are poor. The focal and monotonicity axioms are catered for by the poverty gap ratio and is additively decomposable. Poverty gap ratio is also called income gap ratio. The Poverty gap index is given by;

$$W_1 = \frac{1}{m} \sum_{j=1}^{h} \frac{k - z_j}{k} \tag{3.6}$$

3.2.3 Squared Poverty Gap Ratio

The squared poverty gap ratio has been used by several researchers as a measure of disparities among the poor. This is the aggregate of weights of poverty gaps (as a proportion of poverty line), where the weights are proportions of the poverty gaps themselves. That is a poverty gap of 10% of the poverty line is based on a weight of 10% as compared to a poverty gap ratio where they are weighted equally. Hence, by squaring the poverty gap ratio, more weight is place on individuals below the poverty line. This is written as;

$$W_2 = \frac{1}{m} \sum_{j=1}^{m} \left(\frac{k - z_j}{k} \right)^2 \tag{3.7}$$

3.2.4 Poverty Lines

The quantity of consumption that separates those who are poor from the entire population is called the poverty line. The poverty line examines the quantity of

University of Ghana http://ugspace.ug.edu.gh

consumption below which an individual will be classified as poor. Classification of consumption can be grouped into two parts: those who consume food and those who don't consume food. Hence, the poorer the people, the greater the proportion of the total consumption that distinguishes those who are poor from those who are not poor.

3.3 Measures of Income Inequality

Several researches in economics have been made on measurement of income disparities among household or individual (Atkinson, 1970; Sen, 1973; Anand, 1983; Foster and Sen, 1997). Cowell (2000) referred to income inequality as a disparity in the distribution of income among individuals or households of a particular population in a country. Although there are several ways of assessing income disparities such as variance, coefficient of variation or Robin Hood index (Atkinson, 1969), a common way of measuring inequality is the Gini coefficient which is based on the Lorenz curve (Almas, 2004).

3.3.1 Lorenz Curve

The Lorenz curve generated by Lorenz (1905) is a curve that graphically shows the cumulative distribution function of the probability distribution of income. The Lorenz curve measures the distribution of individual or household income which is used to estimate the various measures of inequality such as the Gini coefficient (Heshmati, 2004). To expantiate, we take note of what was verified by Anand (1997) concerning the distribution of income. The distribution of income can be written as a vector, $z = (z_1, z_2, z_3, ..., z_m)$ where z_j represent individual income in a population with $j = 1, 2, .., m$ and m is the total number

University of Ghana http://ugspace.ug.edu.gh

of individuals. The mean and density function of the income distribution can be written as $\mu(z)$ and $f(z)$.

Income, which is a continuous random variable starts from zero to infinity. i.e $(z_0 \geq 0)$. The curve represents the association that exist between the cumulative proportion of income and the cumulative proportion of individuals that receive income. Let $\pi(z)$ denote the cumulative proportion of the individuals that receive income up to z and $\eta(z)$ represent the cumulative proportion of the total income that is received by individuals in the same population.

The Lorenz curve is a graph that shows the association between π and η. The curve is graphically displayed in a unit square. The 45^0 line joining the points $(0, 0)$ and $(1, 1)$ on the graph is called the egalitarian line. This is because the linear equation $\eta = \pi$ implies that every individual receives equal income. Figure 3.1 below shows that the Lorenz curve falls below the 45^0 line. The Lorenz curve is established by letting z be the income of every individual and $f(z)$ be the probability density function of z . Hence ;

$$\pi(z) = \int_0^z f(x)dx \tag{3.8}$$

and

$$\eta(z) = \frac{1}{\mu}\int_0^z xf(x)dx \tag{3.9}$$

The mean of the distribution of income is ;

$$\mu = \int_0^\infty xf(x)dx \tag{3.10}$$

The graph of $\eta(z)$ against $\pi(z)$ is the Lorenz (1905) curve representing the inequality of income distribution. The deviation of a Lorenz curve from the

University of Ghana http://ugspace.ug.edu.gh

45^0 line for a given income distribution can be measured by several index of income inequality. A perfectly equal individual or household income distribution is always one which means every individual in the population has the same income. The Gini coefficient is the most commonly used measure of income inequality. Figure 3.1 shows the distribution and disparity of income.

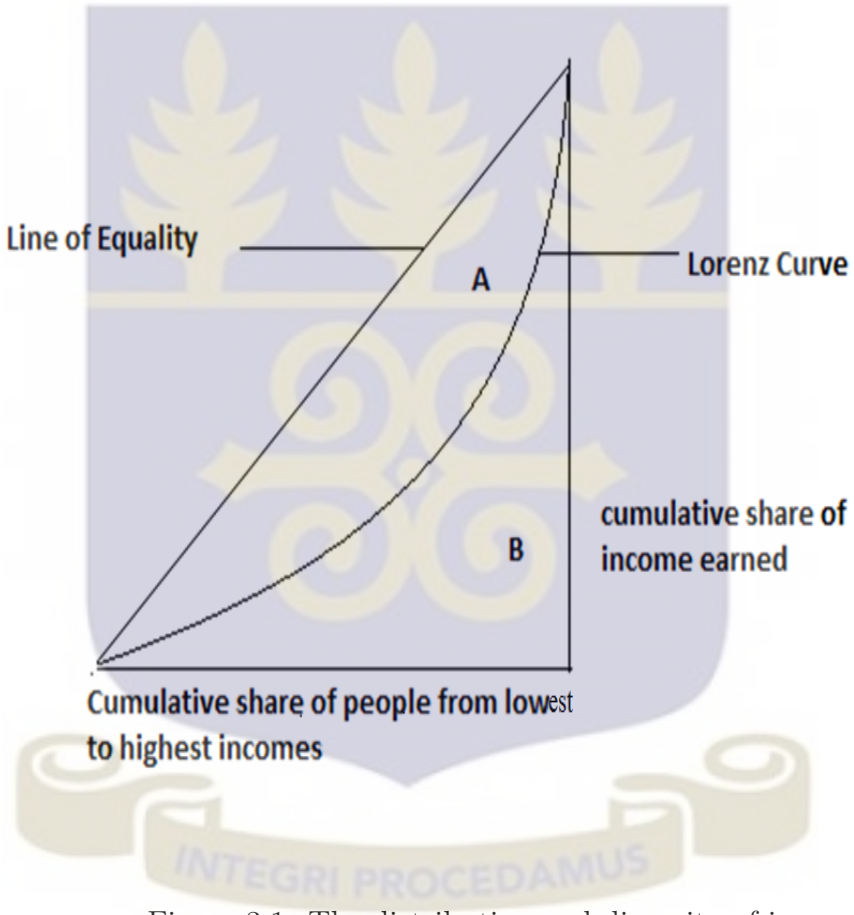


Figure 3.1: The distribution and disparity of income

Ryu & Slottje (1999) used two parametric ways to estimate the Lorenz curve. Firstly, income distribution based on hypothetical statements was used to estimate the Lorenz curve (McDonald & Xu, 1995) and secondly, an estimate of the Lorenz curve was derived by fitting a function to the Lorenz curve (Ryu & Slottje 1996; Chotikapanich & Griffiths 2002). That is, representing z by

University of Ghana http://ugspace.ug.edu.gh

values, the Lorenz curve can be derived by generating the values of $\pi(z)$ and $\eta(z)$ from (3.10) and (3.11). Among the different functional forms proposed for estimating the Lorenz curve, we choose a polynomial proposed by Becker and Weispfenning (1993) and a function proposed by Rasche et.al.(1980). Becker and Weispfenning (1993) 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;

$$\eta(\pi) = a_m\pi^m + a_{m-1}\pi^{m-1} + ... + a_1\pi + a_0 \tag{3.11}$$

When $a_m \neq 0$ and $m \geq 2$, the polynomial function is a continuous non-linear function. Also, estimation of the Lorenz curve proposed by Rasche et.al (1980), with the following nonlinear form is given by;

$$\eta = [1 - (1 - \pi)^s]^{\frac{1}{t}} \tag{3.12}$$

Where η and π being the parameters. This Lorenz curve is not symmetric about the 45^0 line perpendicular to the egalitarian line. Generally, the following conditions must be met by the Lorenz curve (Kakwani and Podder, 1973):

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

University of Ghana http://ugspace.ug.edu.gh

The first condition eliminate the chances of an individual receiving zero or negative income. The fourth condition shows that the curve falls below the 45^0 line. The function of the Lorenz curve is always convex and has a constant slope to ensure that the Lorenz curve always fall in the lower triangle of the unit square (Rasshe et.al.,1980). The estimate of the Lorenz curve is given by;

$$L(\pi) = \int_0^1 \eta(\pi) d\pi \tag{3.13}$$

3.3.2 GINI COEFFICIENT

The Gini coefficient is also referred to as the Gini ratio or Gini index. It is the most frequently used measure of inequality. Statistically, it measures the variation that denote the distribution of a nation's residential income. An Italian statistician and sociologist called Corrado Gini generated it in 1912 and published it in his paper "Variability and Mutability". The Gini coefficient assesses the disparity among income distributions. A Gini coefficient of zero expresses perfect equality, where all individuals receive the same income and a Gini coefficient of one (or100%) shows maximum inequality among individuals where an individual receives all the income.

The Market economy of various countries developed standard measures to describe the income gap. They established that when Gini is less than 0.2, there is absolute equality in the distribution of income; when Gini is between 0.2 and 0.3, the equality in the distribution of income is relatively average; when Gini is between 0.3 and 0.4, the distribution of income is reasonable; when Gini is between 0.4 and 0.5, the income gap is moderate and when Gini is greater than 0.5, the income gap in a country is large (Xiao, Feng & Shiqiang,

University of Ghana http://ugspace.ug.edu.gh

2014). A Gini coefficient greater than one means that some individuals gave negative contributions to the total income. The Gini coefficient is based on the Lorenz curve which is a graph that accounts for the cumulative proportion of household income to the cumulative proportion of income received when the households are put in ascending order of their income. From Figure 3.1 above, when A is the region between the egalitarian line and Lorenz curve and B is the region below the Lorenz curve, then the Gini coefficient is;

$$\frac{A}{A+B} \tag{3.14}$$

When $A = 0$, the Gini coefficient is 0 which explains complete equality, whereas if $B = 0$, the Gini coefficient becomes 1 which explains complete inequality. Since $A + B = 0.5$, the Gini coefficient $G = 2A = 1 - 2B$. If the function $\eta(\pi)$ represent the Lorenz curve, then the Gini coefficient is given by;

$$\text{Gini} = 1 - 2\int_0^1 \eta(\pi)d\pi \tag{3.15}$$

3.4 Quantile-Quantile (Q-Q) Plot

The one variable quantile-quantile (Q-Q) plot is frequently used to examine whether two independent distributions are the same or different (Gnanadesikan & Wilk ,1968). The quantile-quantile (Q-Q) plot is a characteristic graph of the observed data values plotted against the quantiles of the theoretical distribution. If the stipulated theoretical distribution is the exact model, the data points plotted should be in a straight line. The Q-Q plot shows the nature of the tails of the proposed distribution such that if the data points lie along the straight line it implies that the data doesn't deviate from the theoreti-

University of Ghana http://ugspace.ug.edu.gh

cal distribution. The Q-Q plot is formed using the cumulative distribution function, $F(x)$ of the specified theoretical model or distribution. The sample values are arranged from the smallest to the largest and are represented as $Z_{(1)}, Z_{(2)}, Z_{(3)}, .., Z_{(m)}$.For $j = 1, 2, ..., m$ where $Z_j's$ are plotted against the inverse cumulative distribution function;

$$G^{-1}\left[\frac{j - 1/2}{m}\right] \tag{3.16}$$

The utilization of Q-Q in this study is to examine the goodness of fit of income data to the lognormal distribution.

3.5 Kolmogorov-Smirnov Goodness-of-Fit Test

A non-parametric test used to examine whether a sample data belongs to a particular population distribution is the Kolmogorov-Smirnov (K-S) goodness of fit test. The Kolmogorov-Smirnov test depend on the cumulative distribution function of the observed data. It measures the highest distance between the theoretical cumulative distribution function and the distribution function based on the data points. Let $S(x)$ be the empirical distribution function of X which represent the proportion of $X_i's$ that are less than or equal to X for every $X_i, -\infty < X_i < \infty$,

$$S(x) = \frac{1}{m}\sum_{i=1}^{m} I_{\{x_i<x\}}$$

Where $I_{\{x_i\leq x\}}(x)$ is an indicator function and

$$I_{\{x_i\leq x\}}(x) = \{^{0, x_i>x}_{1, x_i\leq x} \tag{3.17}$$

University of Ghana http://ugspace.ug.edu.gh

The cumulative distribution function $F(x)$ is estimated by empirical distribution function $S(x)$. To examine the existence of goodness of fit, we compare the empirical distribution function $S(x)$ with the distribution function based on the hypothesis stated,$F_0(x)$. Conover (1999) proposed that the test statistic is the highest (represented as "sup" for Supremum) vertical distance between $S(x)$ and $F_0(x)$. This is written as;

$$T = Sup\,\|S(x_i) - F_0(x_i)\| \tag{3.18}$$

Testing the hypothesis,

$H_0 : F(x) = F_0(x)$ for all x from $-\infty < x < \infty$

$H_1 : F(x) \neq F_0(x)$ for at least one value of x

The importance of the Kolmogorov-Smirnov test in this study is to test the goodness of fit of the income data to the lognormal distribution.

3.6 Probability Distribution of income

A number of alternative probability density functions have been considered as models of income distribution (McDonald, 1984). These probability distributions include, among others, the exponential, lognormal, gamma, beta, Singh-Maddala, Dagum, Pareto and Weibull distributions. Literature has investigated the extreme values or whole range of the income distribution (Cowell, 1997). The Lognormal and Pareto distributions are frequently used as models of income distributions but Cramer (1971) confirmed that the lognormal distribution fits the income data better. The lognormal distribution used frequently in statistics and probability theory is a continuous distribution function whose logarithm of the observed data points follows the normal

University of Ghana http://ugspace.ug.edu.gh

distribution. Given the random variable X following a distribution which is lognormal, then $Y = In(X)$ will be a random variable that is normally distributed. Also, if Y is normally distributed, then $X = exp(Y)$ will follow a lognormal distribution. The random variable of the lognormal distribution uses values from zero to infinity. The distribution is associated with several names such as the Galton, McAlister, Gibrat and Cobb–Douglas distribution (Norman & Balakrishnan, 1994). Given a normal distribution with mean, μ and standard deviation, σ , then the lognormal probability density function is written as;

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(Inx-\mu)^2}{2\sigma^2}}, x > 0, \sigma > 0, \mu > 0 \tag{3.19}$$

and the cumulative distribution function is;

$$F(x) = \Phi\left(\frac{Inx - \mu}{\sigma}\right) \tag{3.20}$$

A random variable X which follows the lognormal distribution has mean ;

$$E(x) = e^{\mu+\sigma^2/2} \tag{3.21}$$

The variance of X is given by;

$$V(x) = \left(e^{\sigma^2} - 1\right) e^{2\mu+\sigma^2} \tag{3.22}$$

University of Ghana http://ugspace.ug.edu.gh

3.7 Numerical Integration Approach to Find the Area under the Lorenz Curve

Numerical integration involves the approximation of numerical values that cannot be integrated analytically (Gordon, 1998). It is sometimes referred to as quadrature which involves replacing the area under a curve by an area of a square. Several numerical integration methods such as Newton-cotes, Weddle's rule, Romberg integration, Gauss Quadrature and Monte Carlo integration are used to evaluate those functions that can't be integrated analytically. Newton-Cotes methods use interpolating polynomials. Newton-Cotes methods such as the 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 $6th$ order polynomials used respectively.

For the subintervals, the trapezium rule and the Proposed numerical integration method has no restriction but Simpson 1/3 rule must be even, Simpson 3/8 rule must be a multiple of 3, Boole's rule must be a multiple of 4 and Weddle's rule must also be a multiple of 6. Numerical integration has been employed in various fields of science such as biostatistics to estimate quantiles and various distribution functions. It has increased greatly in estimating likelihoods and posterior distributions using Bayesian methods (Evans & Swartz, 1995).

The Newton-Cotes formula is a frequently used interpolator function in the form of a polynomial. This formula involves n points in the interval $[a, b]$ with $n-1$ order polynomial which passes through x_i points and are equally spaced. The Closed Newton-Cotes formula employs Lagrange interpolating when fitting polynomials. Letting $x_0 = a$, $x_n = b$ and $\triangle = \frac{(b-a)}{n}$, Hence;

University of Ghana http://ugspace.ug.edu.gh

$$f(x) = \sum_{i=1}^{n} f(x_i)p_i(x)$$

where each $p_i(x)$ is the Lagrange Polynomial. That is;

$$p_i(x) = \prod_{k=0, i\neq k}^{n} \frac{x - x_k}{x_i - x_k},\ i = 0, 1, 2, .., n \tag{3.23}$$

Integrating $f(x)$ over $[a, b]$ and choosing $x_i = a + \frac{(b-a)i}{n}$ we have the Newton-Cotes rule;

$$A_{n+1}(f) = \int_a^b f(x)dx \approx \sum_{i=0}^{n} w_i f(x_i)$$

Where the the weight w_i is determined by;

$$w_i = \int_a^b p_i(x)dx = \int_{x_0}^{x_n} \prod_{k=0, i\neq k}^{n} \frac{x - x_k}{x_i - x_k},\ i = 0, 1, 2, .., n \tag{3.24}$$

When $n = 1$, the Newton Cotes formula gives the trapezoidal rule of;

$$A_2 = \frac{\triangle}{2}\left[f(x_0) + f(x_1)\right] \tag{3.25}$$

When $n = 2$, it gives the simpson $1/3$ rule of ;

$$A_3 = \frac{\triangle}{3}\left[f(x_0) + 4f(x_1) + f(x_2)\right] \tag{3.26}$$

When $n = 3$, it gives the simpson $3/8$ rule of;

$$A_4 = \frac{3\triangle}{8}\left[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)\right] \tag{3.27}$$

University of Ghana http://ugspace.ug.edu.gh

When $n = 4$, it gives the Boole's rule of;

$$A_5 = \frac{2\triangle}{45}\left[7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)\right] \tag{3.28}$$

When $n = 6$, it gives the Weddle's rule of;

$$A_6 = \frac{3\triangle}{10}\left[f(x_0) + 5f(x_1) + f(x_2) + 6f(x_3) + f(x_4) + 5f(x_5) + f(x_6)\right] \tag{3.29}$$

3.7.1 Trapezoidal Rule

The trapezoidal rule is the easiest and frequently used numerical integration method. It uses the linear function to approximate the original function (Gordon, 1998). It derives a formula for a function of complex variables in the same manner as the other numerical integration methods. The trapezium rule approximate the area under the curve $y = f(x)$ from $x = a$ to $x = b$. The region under the Lorenz curve can be estimated using approximations by a first order interpolating polynomial. In Trapezoidal rule, we approximate$[a, b]$ using two equidistant interpolation points $[a, b]$ corresponding to the endpoints of the interval. Given function values at 2 points when $a = x_0, b = x_1$ and $\triangle = b - a$ as $(x_0, f(x_0))\, and\, (x_1, f(x_1))$, we can estimate $f(x)$ using Lagrange polynomial interpolation. We have the area under the Lorenz curve to be;

$$A_2 = f(x_0)w_o + f(x_1)w_1 \tag{3.30}$$

$$A_2 = \int_a^b \left[\frac{(x - x_1)}{(x_0 - x_1)}f(x_0) + \frac{(x - x_0)}{(x_1 - x_0)}f(x_1)\right] dx \tag{3.31}$$

$$A_2 = \frac{\triangle}{2}\left[f(x_o) + f(x_1)\right]$$

$$A_2 = \frac{b - a}{2}\left[f(x_0) + f(x_1)\right]$$

University of Ghana http://ugspace.ug.edu.gh

$$= \frac{b-a}{2}[f(a) + f(b)] \tag{3.32}$$

The Lorenz curve can be derived by interpolating the function $f(x)$ as a linear polynomial at a and b. The trapezium rule is expected to generate the highest errors compared to other numerical integration methods when the interval is large. However, a precise estimation of the trapezium rule is derived by adding the estimates of the trapezium rule over smaller intervals. The interval $[a, b]$ is divided into n equal intervals by the points $a = x_0 \leq x_1 \leq ... \leq x_n = b$. Let $\triangle = \frac{b-a}{n}$ be the size of the interval, then $x_i = a + ih$, $i = 0, 1, ..., n$. Employing the trapezoidal rule to every interval $[x_i, x_{i+1}]$,we generate the composite trapezoidal rule

$$\int_a^b f(x)dx \approx \triangle \left\{ \frac{f(x_0)}{2} + f(x_1) + ... + f(x_{n-1}) + \frac{f(x_n)}{2} \right\}$$

$$\approx \frac{\triangle}{2} \left\{ f(x_0) + f(x_n) + 2\sum_{i=1}^{n-1} f(x_i) \right\} \tag{3.33}$$

3.7.1.1 Error Analysis for the Trapezoid Rule

Approximating the polynomial function can be used to compute the errors generated by the trapezoidal rule. A linear function generate zero error when the second derivative of the function is zero. That is $f''(x) = 0$ for all x. The second derivative measures the variation of a function from a line. Given $f''(x) = k$ for all x in the interval $[0, x_0]$, then beneath the curve will be quadratic. Since $f''(x) = 2A = k$, we can conclude that $A = \frac{k}{2}$. Since $f(0) = 0, c = 0$ and $f(x_0) = y_0$ then

$$y_0 = \frac{k}{2}x_0^2 + Bx_0$$

University of Ghana http://ugspace.ug.edu.gh

where $B = \frac{y_0}{x_0} - \frac{k}{2}x_0$

Hence the equation of the line is;

$$y = \frac{y_0}{x_0}x \tag{3.34}$$

The area under the curve $f(x)$ is;

$$\begin{aligned}
&\int_0^{x_0} \left[\frac{y_0}{x_0}x - \left(\frac{k}{2}x^2 + \left(\frac{y_0}{x_0} - k\frac{x_0}{2}\right)x\right)\right] dx \\
&= \frac{k}{2}\int_0^{x_0} (x_0 x - x^2)dx \\
&= \frac{k}{2}\left[\frac{x_0^3}{2} - \frac{x_0^3}{3}\right] \\
&= \frac{kx_0^3}{12}
\end{aligned} \tag{3.35}$$

The entire length of interval is divided into n smaller portions with every portion equals, $\triangle = \frac{b-a}{n}$. Hence, the maximum error for every portion would be;

$$\frac{k\triangle^3}{12} = \frac{k(b-a)^3}{12n^3}$$

for n of these errors, the sum of the error is;

$$\begin{aligned}
&= \frac{k(b-a)^3}{12n^3}.n \\
&= \frac{k(b-a)^3}{12n^2}
\end{aligned} \tag{3.36}$$

University of Ghana http://ugspace.ug.edu.gh

3.7.2 Simpson's 1/3 Rule

Simpson's 1/3 rule is an extension of Trapezoidal rule where the integrand is approximated by a second order interpolating polynomial. In Simpson's 1/3 rule, we approximate$[a, b]$ using three equidistant interpolation points $[a, {}^{(a+b)}/_2, b]$ corresponding to the endpoints and midpoint of the interval. Given function values at 3 points when $a = x_0, b = x_2, \frac{a+b}{2} = x_1$and $\triangle = \frac{b-a}{2}$ as $(x_0, f(x_0)), (x_1, f(x_1)), \text{and}(x_2, f(x_2))$, we can estimate $f(x)$ using Lagrange polynomial interpolation. We have the area under the curve $f(x)$ to be;

$$A_3 = f(x_0)w_o + f(x_1)w_1 + f(x_2)w_2 \tag{3.37}$$

$$\begin{aligned} A_3 &= \int_a^b [\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) \\ &= +\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2)]dx \end{aligned}$$

$$A_3 \approx \frac{\triangle}{3}\left[f(x_o) + 4f(x_1) + f(x_2)\right]$$

$$A_3 = \frac{b-a}{6}\left[f(x_0) + 4f(x_1) + f(x_2)\right]$$

$$A_3 = \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \tag{3.38}$$

Simpson's 1/3 rule is usually applied to many short intervals. Letting the x_i be for n equal intervals where n is even, and writing $f_i = f(x_i)$, then the composite Simpson 1/3 rule is ;

$$\begin{aligned} \int_a^b f(x)dx &\approx \frac{\triangle}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + ... + 2f(x_{n-2}) \\ &+ 4f(x_{n-1}) + f(x_n)] \end{aligned}$$

University of Ghana http://ugspace.ug.edu.gh

$$= \frac{\triangle}{3}\left[f(x_0) + 4\sum_{i=1_{i=odd}}^{n-1} f(x_i) + 2\sum_{i=2_{i=even}}^{n-2} f(x_i) + f(x_n)\right] \tag{3.39}$$

3.7.2.1 Error Analysis for the Simpson 1/3 Rule

If $f_h(x)$ denotes the result of applying the composite Simpson 1/3 rule to $f(x)$ over the interval $[a, b]$, and if $f(x)$ has a continuous fourth derivative on (a, b), then the true error in a single application of Simpson's 1/3 rule is given by;

$$\int_a^b f(x)dx - f_h(x) = -\frac{(b-a)^5 f^{(4)}(\xi)}{2880} \tag{3.40}$$

$f^{(4)}$ is the fourth derivative of the function $f(x)$ for some $\xi\epsilon\,[a, b]$. Although Simpson's 1/3 rule was derived to integrate quadratic polynomials exactly on each interval, the presence of the fourth derivative in the error term signals that it in fact integrates cubics exactly as well.

$$E_1 = -\frac{(x_2 - x_0)^5 f^{(4)}(\xi)}{2880}, x_0 < \xi < x_2$$

$$= -\frac{\triangle^5 f^{(4)}(\xi)}{90} \tag{3.41}$$

In a multiple segment Simpson's rule, the error is the sum of the errors in each application of Simpson's 1/3 rule. The total error in the multiple segment Simpson's rule is

$$E_t = -\frac{(b-a)^5 \bar{f}^{(4)}}{90n^4} \tag{3.42}$$

where

$$\bar{f^{(4)}} = \sum_{i=1}^{\frac{n}{2}} \frac{f^{(4)}(\xi)i}{n}$$

is the approximate average value of $f^{(4)}$

University of Ghana http://ugspace.ug.edu.gh

3.7.3 Simpson's 3/8 Rule

Simpson's 3/8 rule is an extension of Simpson 1/3 rule where the integrand is approximated by a third order interpolating polynomial. In Simpson's 3/8 rule, we approximate $[a, b]$ using four equidistant interpolation points $[a, {}^{(2a+b)}/_3, {}^{(a+2b)}/_3, b]$. Given function values at 3 points when $a = x_0, b = x_3, \frac{2a+b}{3} = x_1, \frac{a+2b}{3} = x_2$ and $\triangle = \frac{b-a}{3}$ as $(x_0, f(x_0)), (x_1, f(x_1)), (x_2, f(x_2))$ and $(x_3, f(x_3))$,we can estimate $f(x)$ using Lagrange polynomial interpolation. We have the area under the Lorenz curve to be;

$$A_4 = f(x_0)w_o + f(x_1)w_1 + f(x_2)w_2 + f(x_3)w_3 \tag{3.43}$$

$$A_4 = \int_a^b [\frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f(x_0) + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f(x_1)$$
$$+\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f(x_2) + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}f(x_3)]dx$$

$$A_4 = \int_a^b f(x)dx \approx \frac{3\triangle}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)]$$

$$A_4 = \frac{b-a}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)]$$

$$A_4 = \frac{b-a}{8}\left[f(a) + 3f\left(\frac{2a+b}{3}\right) + 3f\left(\frac{a+2b}{3}\right) + f(b)\right] \tag{3.44}$$

Simpson's 3/8 rule is usually applied to many short intervals by letting the x_i be for n equal intervals where n must be a multiple of 3, and writing $f_i = f(x_i)$. Suppose that the interval $[a, b]$ is subdivided into $3n$ subintervals of equal width, $\triangle = \frac{(b-a)}{3n}$, then the composite Simpson 3/8 rule is;

$$\begin{aligned}\int_a^b f(x)dx &\approx \frac{3\triangle}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + 2f(x_3) + ... + 3f(x_{n-2}) \\ &+ 3f(x_{n-1}) + f(x_n)]\end{aligned}$$

University of Ghana http://ugspace.ug.edu.gh

$$= \frac{3\triangle}{8}\left[f(x_0) + 3\sum_{i=1,4,7}^{n-2} f(x_i) + 3\sum_{i=2,5,8}^{n-1} f(x_i) + 2\sum_{i=3,6,9}^{n-3} f(x_i) + f(x_n)\right] \quad (3.45)$$

3.7.3.1 Error Analysis for the Simpson 3/8 Rule

If $f_h(x)$ denotes the result of applying the composite Simpson 3/8 rule to $f(x)$ over the interval $[a, b]$, and if f has a continuous fourth derivative on(a, b), then the true error in a single application of Simpson's 3/8 rule is given by;

$$\int_a^b f(x)dx - f_h(x) = -\frac{(b-a)^5 f^{(4)}(\xi)}{6480} \quad (3.46)$$

$f^{(4)}$ is the fourth derivative of the function $f(x)$ for some $\xi\epsilon\,[a, b]$. Although Simpson's rule was derived to integrate quadratic polynomials exactly on each interval, the presence of the fourth derivative in the error term signals that it in fact integrates cubics exactly as well.

$$E_1 = -\frac{(x_2 - x_0)^5 f^{(4)}(\xi)}{6480}, x_0 < \xi < x_2$$

$$= -\frac{3\triangle^5 f^{(4)}(\xi)}{80} \quad (3.47)$$

In a multiple segment Simpson's rule, the error is the sum of the errors in each application of Simpson's 3/8 rule. The total error in the multiple segment of Simpson's 3/8 rule is;

$$E_t = -\frac{3(b-a)^5 \bar{f^{(4)}}}{80n^4} \quad (3.48)$$

Where $\bar{f^{(4)}} = \sum_{i=1}^{\frac{n}{2}} \frac{f^{(4)}(\xi)i}{n}$ is the approximate average value of $f^{(4)}$

University of Ghana http://ugspace.ug.edu.gh

3.8 Boole's Rule

Boole's rule is an extension of Simpson's 3/8 rule where the integrand is approximated by a fourth order interpolating polynomial. In Boole's rule, we approximate $[a, b]$ using five equidistant interpolation points. Using Lagrange polynomial interpolation to estimate $f(x)$ with $\triangle = \frac{b-a}{4}$, and putting $n = 4$ into Newton's Cotes Methods, we have the area under the curve $f(x)$ to be;

$$A_5 = f(x_0)w_o + f(x_1)w_1 + f(x_2)w_2 + f(x_3)w_3 + f(x_4)w_4 \tag{3.49}$$

$$\begin{aligned} A_5 = \int_a^b [&\frac{(x-x_1)(x-x_2)(x-x_3)(x-x_4)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)(x_0-x_4)} f(x_0) \\ &+ \frac{(x-x_0)(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)(x_1-x_4)} f(x_1) \\ &+ \frac{(x-x_0)(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)(x_2-x_4)} f(x_2) \\ &+ \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)(x_3-x_4)} f(x_3) \\ &+ \frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_0)(x_4-x_1)(x_4-x_2)(x_4-x_3)} f(x_4)]dx \end{aligned}$$

$$A_5 = \frac{2\triangle}{45}\left[7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)\right]. \tag{3.50}$$

Boole's rule is usually applied to many short intervals by letting the x_i be for n equal intervals where n is a multiple of 4, and writing $f_i = f(x_i)$. Suppose that the interval $[a, b]$ is subdivided into $4n$ subintervals of equal width, $\triangle = \frac{(b-a)}{4n}$ then the composite Boole's rule is;

$$\begin{aligned} \int_a^b f(x)dx \quad &\approx \quad \frac{2\triangle}{45}[7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4) + ... \\ &+ \quad 7f(x_{4n-4}) + 32f(x_{4n-3}) + 12f(x_{4n-2}) + 32f(x_{4n-1}) + 7f(x_{4n})] \end{aligned}$$

University of Ghana http://ugspace.ug.edu.gh

$$= \frac{2\triangle}{45}\sum_{k=1}^{n}[7f(x_{4k-4}) + 32f(x_{4k-3}) + 12f(x_{4k-2}) + 32f(x_{4k-1}) + 7f(x_{4k})] \tag{3.51}$$

3.8.1 Error Analysis for Boole's Rule

If $f_h(x)$ denotes the result of applying the composite Boole's rule to $f(x)$ over the interval $[a, b]$, and if f has a continuous sixth derivative on(a, b), then the true error in a single application of Boole's rule is given by;

$$\int_a^b f(x)dx - f_h(x) = -\frac{(b-a)\triangle^6 f^{(vi)}(\xi)}{4725} \tag{3.52}$$

where $f^{(vi)}$ is the sixth derivative of the function $f(x)$ for some $\xi\epsilon\,[a, b]$. Although Boole's rule was derived to integrate fourth order polynomials exactly on each interval, the presence of the sixth derivative in the error term signals that it in fact integrates cubics exactly as well.

3.9 Weddle's Rule

Weddle's rule is an extension of Boole's rule where the integrand is approximated by a sixth order interpolating polynomial. In Weddle's rule, we approximate $[a, b]$ using seven equidistant interpolation points. Using Lagrange polynomial interpolation to estimate $f(x)$ with $\triangle = \frac{b-a}{6}$, and putting $n = 6$ into Newton's Cotes Methods, we have the area under the Lorenz curve to be;

$$\begin{aligned} A_7 &= f(x_0)w_o + f(x_1)w_1 + f(x_2)w_2 + f(x_3)w_3 \\ &+ f(x_4)w_4 + f(x_5)w_5 + f(x_6)w_6 + f(x_7)w_7 \end{aligned} \tag{3.53}$$

$$A_7 = \frac{3\triangle}{10}[f(x_0) + 5f(x_1) + f(x_2) + 6f(x_3) + f(x_4) + 5f(x_5) + f(x_6)] \tag{3.54}$$

University of Ghana http://ugspace.ug.edu.gh

Weddle's rule is usually applied to many short intervals. Letting the x_i be for n equal intervals, and writing $f_i = f(x_i)$. Suppose that the interval $[a, b]$ is subdivided into $6n$ subintervals of equal width, $\triangle = \frac{(b-a)}{6n}$ then the composite Weddle's rule is;

$$\int_a^b f(x)dx = \frac{3\triangle}{10}[f(x_0)+5f(x_1)+f(x_2)+6f(x_3)+f(x_4)+5f(x_5)+f(x_6)+...$$

$$...+5f(x_{6n-5})+f(x_{6n-4})+6f(x_{6n-3})+f(x_{6n-2})+5f(x_{6n-1})+2f(x_{6n-6})+f(x_{6n})]$$

$$\begin{aligned} &= \frac{3\triangle}{10}\sum_{k=1}^{n}[f(x_{6k-6}) + 5f(x_{6k-5}) + f(x_{6k-4}) + 6f(x_{6k-3}) \\ &+ f(x_{6k-2}) + 5f(x_{6k-1}) + f(x_{6k})] \end{aligned} \tag{3.55}$$

3.9.1 Error Analysis for Weddle's Rule

If $f_h(x)$ denotes the result of applying the composite Weddle's rule to $f(x)$ over the interval $[a, b]$, and if f has a continuous sixth derivative on(a, b), then the true error in a single application of Weddle's rule is given by;

$$\int_a^b f(x)dx - f_h(x) = -\frac{(b-a)\triangle^6 f(x_0)}{840} \tag{3.56}$$

3.10 Proposed Numerical Integration formula

This is an extension of the Trapezoidal rule. Suppose the interval $[a, b]$ is subdivided into $n(n\epsilon z^+)$ equal divisions each of width $\triangle = \frac{b-a}{n}$. Define x_i by $x_i = a + i\triangle, i = 0, 1, 2, ..., n$.Then $x_0 = a$ and $x_n = b$.

Let $f_i = f(x_i), i = 0, 1, 2, ..., n$ be the ordinate at $x_i(i = 0, 1, 2, ..., n)$ of the

University of Ghana http://ugspace.ug.edu.gh

function f. Suppose also that the interval $[x_i, x_{i+1}], i = 0, 1, 2, ..., n-1$;is divided into k equispaced points $x_i + \frac{t}{k}\triangle, t = 1, 2, ..., k$: then the corresponding ordinates of f are given by $f_{i+\frac{t}{k}} = f(x_i + \frac{t}{k}\triangle), t = 1, 2, ..., k; i = 1, 2, ..., n-1$. Clearly,when $t = k, x_i + \frac{t}{k}\triangle = x_{i+1}\, and\, f_{i+\frac{t}{k}} = f_{i+1}$.

The area under the curve for the i^{th} strip is estimated as;

$$A_i = \frac{b-a}{2nk}\left\{f_i + 2\sum_{t=1}^{k} f_{i+\frac{t}{k+1}} + f_{i+1}\right\} \tag{3.57}$$

$$A_i = \frac{\triangle}{2k}\left\{f_i + 2\sum_{t=1}^{k} f_{i+\frac{t}{k+1}} + f_{i+1}\right\} \tag{3.58}$$

When $k = 1$,we have

$$A_i = \frac{\triangle}{2}\left[f_i + f_{i+1}\right]$$

When $k = 2$, we have

$$A_i = \frac{\triangle}{4}\left[f_i + 2f_{i+\frac{1}{2}} + f_{i+1}\right]$$

When $k = 3$, we have

$$A_i = \frac{\triangle}{6}\left[f_i + 2\left(f_{i+\frac{1}{3}} + f_{i+\frac{2}{3}}\right) + f_{i+1}\right]$$

PROOF

when $k = 1$, $\triangle = \frac{b-a}{n}$ and $i = 0, 1, .., n-1$

$$A_i = \triangle Min(f_i, f_{i+1}) + \frac{\triangle}{2}\{Max(f_i, f_{i+1}) - Min(f_i, f_{i+1})\}$$

University of Ghana http://ugspace.ug.edu.gh

When f is a monotone increasing over (x_i, x_{i+1}), A_i reduces to

$$A_i = \triangle f_i + \frac{\triangle}{2}(f_{i+1} - f_i)$$

$$A_i = \frac{\triangle}{2}(f_i + f_{i+1})$$

When f is a monotone decreasing over (x_i, x_{i+1}), $A_i$reduces to

$$A_i = \triangle f_{i+1} + \frac{\triangle}{2}(f_i - f_{i+1})$$

$$A_i = \frac{\triangle}{2}(f_i + f_{i+1})$$

This gives the trapezium rule.

when $k = 2$ and $\triangle = \frac{b-a}{n}$, then;

$$\begin{aligned} A_i &= \triangle Min(f_i, f_{i+1}) + \frac{\triangle}{2}\left\{f_{i+\frac{1}{2}} - Min(f_i, f_{i+1})\right\} \\ &+ \frac{\triangle}{4}\left\{f_{i+\frac{1}{2}} - Min(f_i, f_{i+1})\right\} + \frac{\triangle}{4}\left\{Max(f_i, f_{i+1}) - f_{i+\frac{1}{2}}\right\} \end{aligned}$$

$$A_i = \frac{\triangle}{4}Min(f_i, f_{i+1}) + \frac{\triangle}{2}f_{i+\frac{1}{2}} + \frac{\triangle}{4}Max(f_i, f_{i+1})$$

When f is a monotone increasing over (x_i, x_{i+1}), A_i, reduces to

University of Ghana http://ugspace.ug.edu.gh

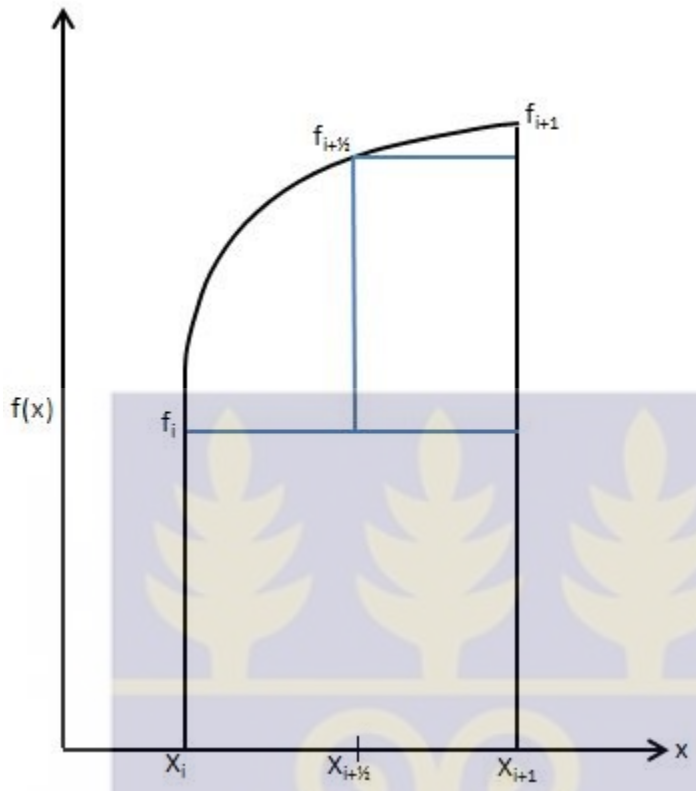


Figure 3.2: Monotonically Increasing function of $f(x)$ when $k = 1$

$$A_i = \frac{\triangle}{4} f_i + \frac{\triangle}{2} f_{i+\frac{1}{2}} + \frac{\triangle}{4} f_{i+1}$$

$$A_i = \frac{\triangle}{4} (f_i + 2f_{i+\frac{1}{2}} + f_{i+1})$$

When f is a monotone decreasing over (x_i, x_{i+1}), A_i, reduces to

University of Ghana http://ugspace.ug.edu.gh

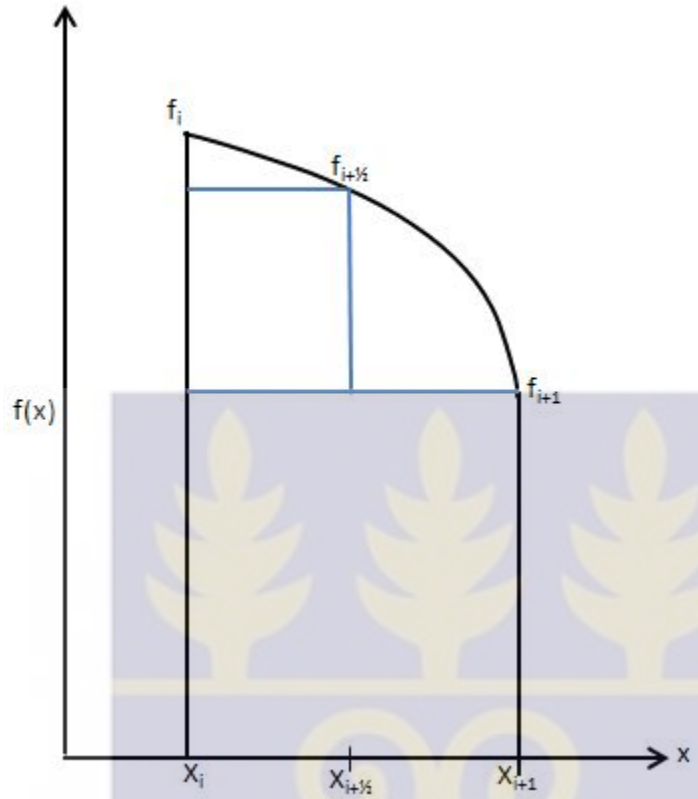


Figure 3.3: Monotonically decreasing function of $f(x)$ when $k = 1$

$$A_i = \frac{\triangle}{4} f_{i+1} + \frac{\triangle}{2} f_{i+\frac{1}{2}} + \frac{\triangle}{4} f_i$$

$$A_i = \frac{\triangle}{4}(f_i + 2f_{i+\frac{1}{2}} + f_{i+1})$$

when $k = 3$ and $\triangle = \frac{b-a}{n}$, then;

$$\begin{aligned} A_i &= \triangle Min(f_i, f_{i+1}) + \frac{\triangle}{6}\left\{f_{i+\frac{1}{3}} - Min(f_i, f_{i+1})\right\} \\ &+ \frac{\triangle}{6}\left\{f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right\} + \frac{\triangle}{6}\left\{Max(f_i, f_{i+1}) - f_{i+\frac{2}{3}}\right\} \\ &+ \frac{2}{3}\triangle\left\{f_{i+\frac{1}{3}} - Min(f_i, f_{i+1})\right\} + \frac{\triangle}{3}\left\{f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right\} \end{aligned}$$

When f is a monotone increasing over (x_i, x_{i+1}), A_i, reduces to

University of Ghana http://ugspace.ug.edu.gh

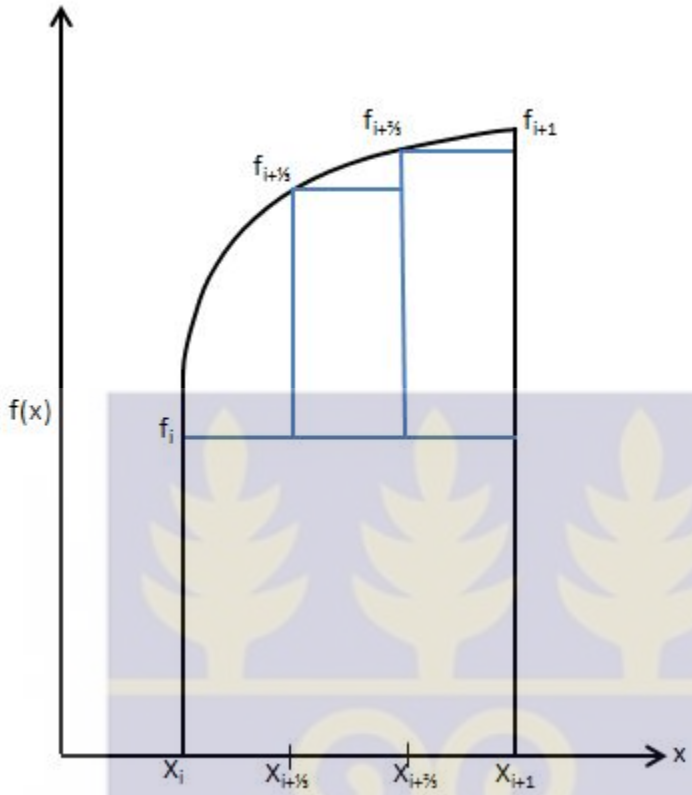


Figure 3.4: Monotonically Increasing function of $f(x)$ when $k = 2$

$$A_i = \triangle f_i + \frac{\triangle}{6}\left(f_{i+\frac{1}{3}} - f_i\right) + \frac{\triangle}{6}\left(f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right) + \frac{\triangle}{6}\left\{f_{i+1} - f_{i+\frac{2}{3}}\right\} + \frac{2\triangle}{3}\left\{f_{i+\frac{1}{3}} - f_i\right\} + \frac{\triangle}{3}\left\{f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right\}$$

$$A_i = \frac{\triangle}{6} f_i + \frac{\triangle}{3} f_{i+\frac{1}{3}} + \frac{\triangle}{3} f_{i+\frac{2}{3}} + \frac{\triangle}{6} f_{i+1}$$

$$A_i = \frac{\triangle}{6}\left[f_i + 2\left(f_{i+\frac{1}{3}} + f_{i+\frac{2}{3}}\right) + f_{i+1}\right]$$

When f is a monotone decreasing over (x_i, x_{i+1}), A_i, reduces to

University of Ghana http://ugspace.ug.edu.gh

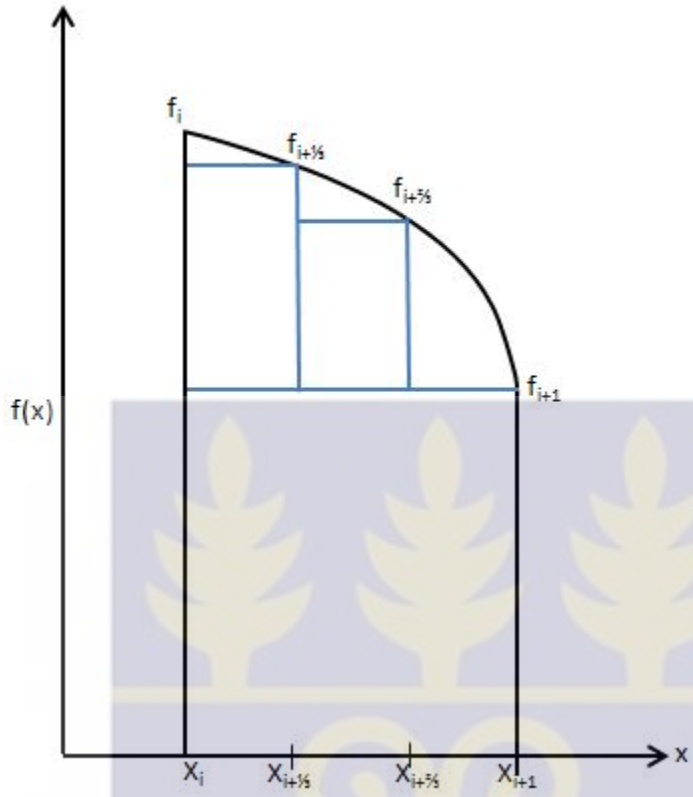


Figure 3.5: Monotonically decreasing function of $f(x)$ when $k = 2$

$$\begin{aligned} A_i &= \triangle f_{i+1} + \frac{\triangle}{6}\left(f_{i+\frac{2}{3}} - f_{i+1}\right) + \frac{\triangle}{6}\left(f_{i+\frac{1}{3}} - f_{i+\frac{2}{3}}\right) + \frac{\triangle}{6}\left\{f_i - f_{i+\frac{1}{3}}\right\} \\ &+ \frac{2\triangle}{3}\left\{f_{i+\frac{2}{3}} - f_{i+1}\right\} + \frac{\triangle}{3}\left\{f_{i+\frac{1}{3}} - f_{i+\frac{2}{3}}\right\} \end{aligned}$$

$$A_i = \frac{\triangle}{6} f_{i+1} + \frac{\triangle}{3} f_{i+\frac{2}{3}} + \frac{\triangle}{3} f_{i+\frac{1}{3}} + \frac{\triangle}{6} f_i$$

$$A_i = \frac{\triangle}{6}\left[f_i + 2\left(f_{i+\frac{1}{3}} + f_{i+\frac{2}{3}}\right) + f_{i+1}\right]$$

The proof of $k \geq 4$ can similarly be obtained. Clearly, the results are the same irrespective of whether is monotone increasing or decreasing; hence result

University of Ghana http://ugspace.ug.edu.gh

3.11 Proposed Composite Numerical Integration Method

The proposed composite method provides a formula for estimating numerically the area under the curve of f and above the horizontal axis between the intervals $[a, b]$. It is the sum of the areas of all the n strips each of width $\triangle = \frac{b-a}{n}$ and k sub-divisions at the top as indicated in the diagram. Thus the proposed composite numerical integration method is given by;

$$A_n = \frac{\triangle}{2k}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + 2\sum_{i=0}^{n-1}\sum_{t=1}^{k-1} f_{i+\frac{t}{k}} + f_n\right] \tag{3.59}$$

Where the subscript in A_n means the area estimation is based on n equispaced intervals in the interval $[a, b]$. The symbols in (67) are as previously defined.

PROOF

The proof follows trivially by summing the $A_i (i = 0, 1, 2, ..., n-1)$ in equation (66).

when $k = 1$ and $\triangle = \frac{b-a}{n}$

$$A_n = \frac{\triangle}{2}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + f_n\right]$$

when $k = 2$ and $\triangle = \frac{b-a}{n}$;

$$A_n = \frac{\triangle}{4}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + 2\sum_{i=0}^{n-1} f_{i+\frac{1}{2}} + f_n\right]$$

when $k = 2$ and $\triangle = \frac{b-a}{n}$;

$$A_n = \frac{\triangle}{6}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + 2\sum_{i-0}^{n-1} f_{i+\frac{1}{3}} + 2\sum_{i=0}^{n-1} f_{i+\frac{2}{3}} + f_n\right]$$

University of Ghana http://ugspace.ug.edu.gh

PROOF

when $k = 1$ and $\triangle = \frac{b-a}{n}$

$$\begin{aligned} A_n &= \triangle Min(f_0, f_1) + \frac{\triangle}{2}\{Max(f_0, f_1) - Min(f_0, f_1)\} \\ &+ \triangle Min(f_1, f_2) + \frac{\triangle}{2}\{Max(f_1, f_2) - Min(f_1, f_2)\} \\ &+ .... + \triangle Min(f_{n-1}, f_n) + \frac{\triangle}{2}\{Max(f_{n-1}, f_n) - Min(f_{n-1}, f_n)\} \end{aligned}$$

$$A_n = \triangle \sum_{i=0}^{n-1} \left[Min(f_i, f_{i+1}) + \frac{1}{2}\{Max(f_i, f_{i+1}) - Min(f_i, f_{i+1})\} \right]$$

When $f_i = f(x_i)$ is a monotone increasing over (x_i, x_{i+1}), A_n, reduces to

$$A_n = \triangle \sum_{i=0}^{n-1} \left[f_i + \frac{1}{2}(f_{i+1} - f_i) \right]$$

$$A_n = \frac{\triangle}{2} \sum_{i=0}^{n-1} [f_i + f_{i+1}]$$

$$A_n = \frac{\triangle}{2} \left[f_0 + 2\sum_{i=1}^{n-1} f_i + f_n \right]$$

When $f_i = f(x_i)$ is a monotone decreasing over (x_i, x_{i+1}), A_n, reduces to

$$A_n = \triangle \sum_{i=0}^{n-1} \left[f_{i+1} + \frac{1}{2}(f_i - f_{i+1}) \right]$$

$$A_n = \frac{\triangle}{2} \sum_{i=0}^{n-1} [f_{i+1} + f_i]$$

$$A_n = \frac{\triangle}{2} \left[f_0 + 2\sum_{i=1}^{n-1} f_i + f_n \right]$$

This gives the trapezium rule.

University of Ghana http://ugspace.ug.edu.gh

when $k = 2$ and $\triangle = \frac{\text{b}-\text{a}}{\text{n}}$

$$\begin{aligned} A_n &= \triangle Min(f_0, f_1) + \frac{\triangle}{4}\left\{f_{\frac{1}{2}} - Min(f_0, f_1)\right\} \\ &+ \frac{\triangle}{2}\left\{f_{\frac{1}{2}} - Min(f_0, f_1)\right\} + \frac{\triangle}{4}\left\{Max(f_0, f_1) - f_{\frac{1}{2}}\right\} + ... \\ &+ \triangle Min(f_{n-1}, f_n) + \frac{\triangle}{4}\left\{f_{n-\frac{1}{2}} - Min(f_{n-1}, f_n)\right\} \\ &+ \frac{\triangle}{2}\left\{f_{n-\frac{1}{2}} - Min(f_{n-1}, f_n)\right\} + \frac{\triangle}{4}\left\{Max(f_{n-1}, f_n) - f_{n-\frac{1}{2}}\right\} \end{aligned}$$

$$\begin{aligned} A_n &= \triangle \sum_{i=0}^{n-1}[Min(f_i, f_{i+1}) + \frac{1}{4}\left\{f_{i+\frac{1}{2}} - Min(f_i, f_{i+1})\right\} \\ &+ \frac{1}{2}\left\{f_{i+\frac{1}{2}} - Min(f_i, f_{i+1})\right\} - \frac{1}{4}\left\{Max(f_i, f_{i+1}) - f_{i+\frac{1}{2}}\right\}] \end{aligned}$$

when $f_i = f(x_i)$ is a monotone increasing over (x_i, x_{i+1}), A_n, reduces to

$$A_n = \triangle \sum_{i=0}^{n-1}\left[f_i + \frac{1}{4}\left(f_{i+\frac{1}{2}} - f_i\right) + \frac{1}{2}\left(f_{i+\frac{1}{2}} - f_i\right) + \frac{1}{4}\left(f_{i+1} - f_{i+\frac{1}{2}}\right)\right]$$

$$A_n = \frac{\triangle}{4}\sum_{i=0}^{n-1}\left[f_i + 2f_{i+\frac{1}{2}} + f_{i+1}\right]$$

$$A_n = \frac{\triangle}{4}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + 2\sum_{i=0}^{n-1} f_{i+\frac{1}{2}} + f_n\right]$$

when $f_i = f(x_i)$ is a monotone decreasing over (x_i, x_{i+1}), A_n, reduces to

$$A_n = \triangle \sum_{i=0}^{n-1}\left[f_{i+1} + \frac{1}{4}\left(f_{i+\frac{1}{2}} - f_{i+1}\right) + \frac{1}{2}\left(f_{i+\frac{1}{2}} - f_{i+1}\right) + \frac{1}{4}\left(f_i - f_{i+\frac{1}{2}}\right)\right]$$

$$A_n = \frac{\triangle}{4}\sum_{i=0}^{n-1}\left[f_{i+1} + 2f_{i+\frac{1}{2}} + f_i\right]$$

University of Ghana http://ugspace.ug.edu.gh

$$A_n = \frac{\triangle}{4}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + 2\sum_{i=0}^{n-1} f_{i+\frac{1}{2}} + f_n\right]$$

when $k = 3$ and $\triangle = \frac{\text{b-a}}{\text{n}}$, then;

$$\begin{aligned}
A_n &= \triangle Min(f_0, f_1) + \frac{\triangle}{6}\left\{f_{\frac{1}{3}} - Min(f_0, f_1)\right\} \\
&+ \frac{\triangle}{6}\left\{f_{\frac{2}{3}} - f_{\frac{1}{3}}\right\} + \frac{\triangle}{6}\left\{Max(f_0, f_1) - f_{\frac{2}{3}}\right\} \\
&+ \frac{2\triangle}{3}\left\{f_{\frac{1}{3}} - Min(f_0, f_1)\right\} + \frac{\triangle}{3}\left\{f_{\frac{2}{3}} - f_{\frac{1}{3}}\right\} + ... \\
&+ \triangle Min(f_{n-1}, f_n) + \frac{\triangle}{6}\left\{f_{n-\frac{1}{3}} - Min(f_{n-1}, f_n)\right\} \\
&+ \frac{\triangle}{6}\left\{f_{n-\frac{2}{3}} - f_{n-\frac{1}{3}}\right\} + \frac{\triangle}{6}\left\{Max(f_{n-1}, f_n) - f_{n-\frac{2}{3}}\right\} \\
&+ \frac{2\triangle}{3}\left\{f_{n-\frac{1}{3}} - Min(f_{n-1}, f_n)\right\} + \frac{\triangle}{3}\left\{f_{n-\frac{2}{3}} - f_{n-\frac{1}{3}}\right\}
\end{aligned}$$

$$\begin{aligned}
A_n &= \triangle\sum_{i=0}^{n}[Min(f_i, f_{i+1}) + \frac{1}{6}\left\{f_{i+\frac{1}{3}} - Min(f_i, f_{i+1})\right\} \\
&+ \frac{1}{6}\left\{f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right\} + \frac{1}{6}\left\{Max(f_i, f_{i+1}) - f_{i+\frac{2}{3}}\right\} \\
&+ \frac{2}{3}\left\{f_{i+\frac{1}{3}} - Min(f_i, f_{i+1})\right\} + \frac{1}{3}\left\{f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right\}]
\end{aligned}$$

when f is a monotone increasing over (x_i, x_{i+1}), A_n, reduces to

$$\begin{aligned}
A_n &= \triangle\sum_{i=0}^{n}[f_i + \frac{1}{6}\left(f_{i+\frac{1}{3}} - f_i\right) + \frac{1}{6}\left(f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right) + \frac{1}{6}\left\{f_{i+1} - f_{i+\frac{2}{3}}\right\} \\
&+ \frac{2}{3}\left\{f_{i+\frac{1}{3}} - f_i\right\} + \frac{1}{3}\left\{f_{i+\frac{2}{3}} - f_{i+\frac{1}{3}}\right\}]
\end{aligned}$$

$$A_n = \frac{\triangle}{6}\sum_{i=0}^{n-1}\left[f_i + 2\left(f_{i+\frac{1}{3}} + f_{i+\frac{2}{3}}\right) + f_{i+1}\right]$$

University of Ghana http://ugspace.ug.edu.gh

$$A_n = \frac{\triangle}{6}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + 2\sum_{i-0}^{n-1} f_{i+\frac{1}{3}} + 2\sum_{i=0}^{n-1} f_{i+\frac{2}{3}} + f_n\right]$$

when f is a monotone decreasing over (x_i, x_{i+1}), A_n, reduces to

$$\begin{aligned} A_n &= \triangle\sum_{i=0}^{n}[f_{i+1} + \frac{1}{6}\left(f_{i+\frac{2}{3}} - f_{i+1}\right) + \frac{1}{6}\left(f_{i+\frac{1}{3}} - f_{i+\frac{2}{3}}\right) + \frac{1}{6}\left\{f_i - f_{i+\frac{1}{3}}\right\} \\ &+ \frac{2}{3}\left\{f_{i+\frac{2}{3}} - f_{i+1}\right\} + \frac{1}{3}\left\{f_{i+\frac{1}{3}} - f_{i+\frac{2}{3}}\right\}] \end{aligned}$$

$$A_n = \frac{\triangle}{6}\sum_{i=0}^{n-1}\left[f_i + 2\left(f_{i+\frac{1}{3}} + f_{i+\frac{2}{3}}\right) + f_{i+1}\right]$$

$$A_n = \frac{\triangle}{6}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + 2\sum_{i-0}^{n-1} f_{i+\frac{1}{3}} + 2\sum_{i=0}^{n-1} f_{i+\frac{2}{3}} + f_n\right]$$

The cases for $k \geq 4$ can similarly be generated

EXAMPLE

Find $\int_0^2 e^{x^2} dx$. Taking 12 intervals

Solution

Method 1: Exact (Numerical Integration)

$$A = \int_0^2 e^{x^2} = \int_0^2 \sum_{n=0}^{\infty} \frac{x^{2n}}{n!} dx$$

$$= \sum_{n=0}^{\infty} \int_0^2 \frac{x^{2n}}{n!} dx$$

$$= \sum_{n=0}^{\infty} \frac{2^{2n+1}}{(2n+1)n!}$$

$$= 2 + \frac{2^3}{3} + \frac{2^5}{5(2!)} + \frac{2^7}{7(3!)} + ... + \frac{2^{25}}{25(12!)} + ...$$

University of Ghana http://ugspace.ug.edu.gh

$= 16.45262777$

Method 2: Trapezium rule

Using the trapezium rule, let $n = 12$ be the number of intervals.

$\triangle = \frac{b-a}{n} = \frac{2-0}{12} = 0.16667$. Let $y = f(x)$, where $y = e^{x^2}$

x	0	0.1667	0.3333	0.5	0.6667	0.8333	1.0
y	1	1.0282	1.1175	1.2840	1.5596	2.0026	2.7183

x	1.1667	1.3333	1.5	1.6667	1.8333	2
y	3.9005	5.9167	9.4877	16.0832	28.8212	54.5982

$A = \frac{0.1667}{2}[1 + 2(1.0282 + 1.1175 + ... + 28.8212) + 54.5982]$

$= \frac{1}{12}[1 + 2(73.9196) + 54.5982]$

$= \frac{203.4374}{12} = 16.95311$

Method 3: Simpson's 1/3 rule

Using the Simpson's 1/3 rule, let $n = 12$ be the number of intervals.

$\triangle = \frac{b-a}{n} = \frac{2-0}{12} = 0.1667$. Let $y = f(x)$,where $y = e^{x^2}$

x	0	0.1667	0.3333	0.5	0.6667	0.8333	1.0
y	1	1.0282	1.1175	1.2840	1.5596	2.0026	2.7183

x	1.1667	1.3333	1.5	1.6667	1.8333	2
y	3.9005	5.9167	9.4877	16.0832	28.8212	54.5982

$A = \frac{0.1667}{3}[1 + 4(1.0282 + 1.2840 + ... + 28.8212) +$

$2(1.1175 + 1.5596 + ... + 16.0832) + 54.5982$

$= \frac{1}{18}[1 + 4(46.5243) + 2(27.3954) + 54.5982]$

$= \frac{296.4859}{18} = 16.47144$

Method 4: Simpson's 3/8 rule

University of Ghana http://ugspace.ug.edu.gh

Using the Simpson's 3/8 rule, let $n = 12$ be the number of intervals.

$\triangle = \frac{b-a}{n} = \frac{2-0}{12} = 0.1667$. Let $y = f(x)$, where $y = e^{x^2}$

x	0	0.1667	0.3333	0.5	0.6667	0.8333	1.0
y	1	1.0282	1.1175	1.2840	1.5596	2.0026	2.7183

x	1.1667	1.3333	1.5	1.6667	1.8333	2
y	3.9005	5.9167	9.4877	16.0832	28.8212	54.5982

$$A = \frac{3(0.1667)}{8}[1 + 3(1.0282 + 1.1175 + 28.8212) + 2(1.2840 + 2.7183 + 9.4877) + 54.5982]$$

$$= \frac{1}{16}[1 + 3(60.4296) + 2(13.4900) + 54.5982]$$

$$= \frac{263.8669}{16} = 16.4917$$

Method 5: Boole's rule

Using the Boole's rule, let $n = 12$ be the number of intervals.

$\triangle = \frac{b-a}{n} = \frac{2-0}{12} = 0.1667$. Let $y = f(x)$, where $y = e^{x^2}$

x	0	0.1667	0.3333	0.5	0.6667	0.8333	1.0
y	1	1.0282	1.1175	1.2840	1.5596	2.0026	2.7183

x	1.1667	1.3333	1.5	1.6667	1.8333	2
y	3.9005	5.9167	9.4877	16.0832	28.8212	54.5982

$$A = \frac{2(0.1667)}{45}[7(1 + 54.5982) + 32(1.0282 + 1.2840 + 2.0026 + ... + 28.8212) + 12(1.1175 + 2.7183 + 16.0832) + 14(1.5596 + 5.9167)]$$

$$= \frac{1}{135}[7(55.5982) + 32(46.5243) + 12(19.91904) + 14(7.4763)]$$

$$= \frac{2221.66}{135} = 16.4567$$

Method 6: Weddle's rule

University of Ghana http://ugspace.ug.edu.gh

Using the Weddle's rule, let $n = 12$ be the number of intervals.

$\triangle = \frac{b-a}{n} = \frac{2-0}{12} = 0.1667$. Let $y = f(x)$, where $y = e^{x^2}$

x	0	0.1667	0.3333	0.5	0.6667	0.8333	1.0
y	1	1.0282	1.1175	1.2840	1.5596	2.0026	2.7183

x	1.1667	1.3333	1.5	1.6667	1.8333	2
y	3.9005	5.9167	9.4877	16.0832	28.8212	54.5982

$$A = \frac{3(0.1667)}{10}[1 + 5(1.0282 + 3.9005) + (1.1175 + 5.9167) + 6(1.2840 + 9.4877) + (1.5596 + 16.0832) + 5(2.0026 + 28.8212) + 2(2.7183) + 54.5982]$$

$$= \frac{1}{20}[1 + 5(33.0187) + 6(10.7717) + 2(2.7183) + 24.677 + 54.5982]$$

$$= \frac{329.1048}{20} = 16.4552$$

Method 7: Proposed Numerical Integration Method

Using the Proposed Numerical Integration Method, Let $n = 12$ be the number of intervals.

$\triangle = \frac{b-a}{n} = \frac{2-0}{12} = 0.16667$. Let $y = f(x)$, where $y = e^{x^2}$

when $k = 1$,we get the exact estimate as the Trapezium rule

when k=2;

x	0	0.0833	0.1667	0.25	0.3333	0.4167	0.5
y	1	1.0070	1.0282	1.0645	1.1175	1.1896	1.2840

x	0.5833	0.6667	0.75	0.8333	0.9167	1
y	1.4053	1.5596	1.7551	2.0026	2.3170	2.7182

x	1.0833	1.1667	1.25	1.3333	1.4167	1.5
y	3.2336	3.9005	4.7707	5.9167	7.4405	9.4877

University of Ghana http://ugspace.ug.edu.gh

x	1.5833	1.6667	1.275	1.8333	1.9167	2
y	12.2674	16.0832	21.3809	28.8212	39.3939	54.5982

$A = \frac{0.1667}{4}\left[1 + 2(1.0070 + 1.0282 + 1.0645 + 1.1175 + ... + 39.3939) + 54.5982\right]$

$= \frac{1}{24}\left[1 + 2(171.1452) + 54.5982\right]$

$= \frac{397.8886}{24} = 16.5787$

Generally, when $k = 8, 9, ..., 20$, the values for the Proposed Numerical Integration method for evaluating $\int_0^2 e^{x^2} dx$, taking 12 intervals are given in the table 3.1 below.

Table 3.1: Estimates of the Area using numerical integration method with 12 subintervals

Integration Method	Estimated Area	Error
Exact	16.45263	-
Trapezium rule	16.95311	0.5016
Simpson's 1/3rule	16.47144	0.0199
Simpson's 3/8 rule	16.49168	0.0402
Boole's rule	16.45674	0.0052
Weddle's rule	16.45524	0.0037
Proposed method when k=8	16.46053	0.0079
Proposed method when k=9	16.45887	0.00624
Proposed method when k=10	16.45681	0.00418
Proposed method when k=12	16.45614	0.00351
Proposed method when k=13	16.45562	0.00299
Proposed method when k=14	16.45521	0.00258
Proposed method when k=15	16.45487	0.00224
Proposed method when k=20	16.45389	0.00126

Comparing the Proposed Numerical Integration Method to the various numerical integration formulas from Table 3.1 above using the errors, when $k = 1$, the area under the curve is 16.95311 which is the same as the area under the curve using the Trapezium rule. Also when $k = 2$ and above, the estimates of the area under the curve are better than the Trapezium rule. That is, the Trapezium rule yields the highest error of 0.50048 as compared to the exact

University of Ghana http://ugspace.ug.edu.gh

numerical integration. Furthermore, when k = 8 and above, the estimates of the area under the curve are better than the Simpson1/3rule. When $k = 12$ and above, the estimates of the area under the curve are better than Boole's rule and when k = 13 and above, the estimates of the area under the curve are better than Weddle's rule. The proposed method with smaller divisions of the interval gives a better estimate with lesser errors as compared to the Trapezium rule, Simpson 1/3 and 3/4 rule, Boole's and Weddle's rule. Given an error bound of ≤ 0.003, clearly as the $\triangle$decrease, k decreases. Hence it suffices to determine the relationship between k and $\triangle$ so that for the function in equation 66, one can determine the number of divisions k at the top to arrive at the desired estimate with the stated error bound. To get the equation, we regress to get;

$k = 79.383\triangle$

With a coefficient of determination of 0.9905.

To find the equation, a scatter plot was drawn to find the relationship as shown in Figure 3.2. From the example above, if $n = 20, \triangle = \frac{2-0}{20} = 0.1$, then one need $k = 79.383(0.1) \approx 8$ to get the estimated area with an error bound of less than 0.003.

Further research can be made on several groups of functions to show whether the same relationship exit as the function in the example so that when k is decreasing, $\triangle$ is decreasing.

University of Ghana http://ugspace.ug.edu.gh

Chapter 4

ANALYSIS AND DISCUSSIONS

4.1 Introduction

This chapter basically discusses the income distribution of Ghanaian income data from the Ghana Living Standard Survey 2012/2013.The Q-Q plot which is a diagnostics test is used to graphically demonstrate goodness of fit to the lognormal distribution. The Goodness-of-Fit test (Kolmogorov-Smirnov test) was used to statistically test fitness of the lognormal distribution to the income data. The proposed numerical integration method was used to estimate the Lorenz curve and Gini coefficient of income.

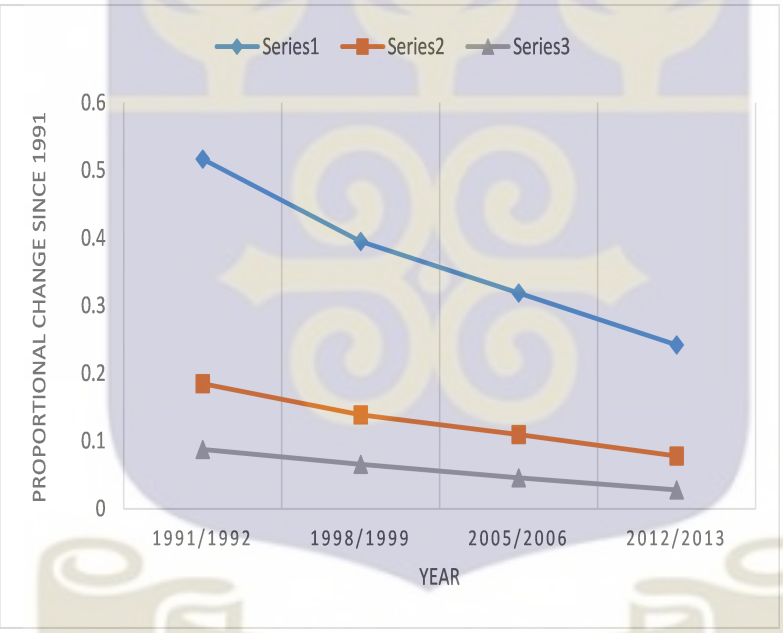
4.2 Empirical trends in Poverty Incidence, Poverty gap and Squared Poverty gap in Ghana from 1991/92 to 2012/2013

This section presents trends of how poverty has changed in Ghana over time since the late 1990s. The work on recent poverty trends is based on the

University of Ghana http://ugspace.ug.edu.gh

2012/2013 nationally representative Ghana Living Standards Survey (GLSS) which is a household survey conducted by Ghana Statistical Services in 2013. The results from the GLSS 5 and GLSS 6 surveys shown in Figure 4.1 indicate that given a poverty line of GH¢1,314, poverty reduced by 7.7 percentage over the seven-year period (2005/2006 to 2012/13).

Since the current (2013) poverty rate is slightly less than half the rate recorded in 1991/92, the estimates presented in Figure 4. 1 suggest that the share of the population living in poverty was reduced from 51.7 percent in 1991/92 to 39.5 percent in 1998/99 and 31.9 percent in 2005/2006 to 24.2 in 2012/2013.



Source:Ghana Living Standard Survey 6

Figure 4.1: Proportional Changes in Poverty Incidence, Poverty Gap and Squared Poverty Gap in Ghana from 1991 to 2013

4.3 Data

The data is derived from the third, fourth, fifth and sixth rounds of the Ghana Living Standards Survey (GLSS) which was collected in 1991/92,

University of Ghana http://ugspace.ug.edu.gh

1998/99, 2005/06, 2012/13. The GLSS is a multi-purpose survey of households in Ghana, which collects information on many different dimensions of living conditions, among others, on education, health and employment. The data is collected on a countrywide basis. Data from the GLSS 6 had a total income of GHC 244,759,213.2 from 58,796 individuals which was used to perform this analysis. It is observed from Table 4.2 that a substantial difference exists between the maximum and minimum income values with mean 14,788.18. This is however not surprising considering the magnitude of the skewness and kurtosis values.

Table 4.1: Descriptive Statistics

GLSS 6 Income	Value
Mean	14788.18
Maximum	2184471
Minimum	2
Standard Deviation	46257.611
Skewness	16.709
Kurtosis	479.105
Number of observations	16551

Figure 4.2 illustrates the histogram of the income data. It can be observed that the empirical distribution function shows a very high peak around the mean and also relatively fatter tail to the right compared to normal. This shows that the data deviates from normality as established above. The histogram shows that relatively more observations lie to the right of the mean of the distribution as compared to the left. This is also in agreement of the positive skewness obtained and the high peak which corresponds to the large kurtosis value obtained. The income is heavy tailed which conformed to the studies in the literature that, incomes are heavy-tailed.

University of Ghana http://ugspace.ug.edu.gh

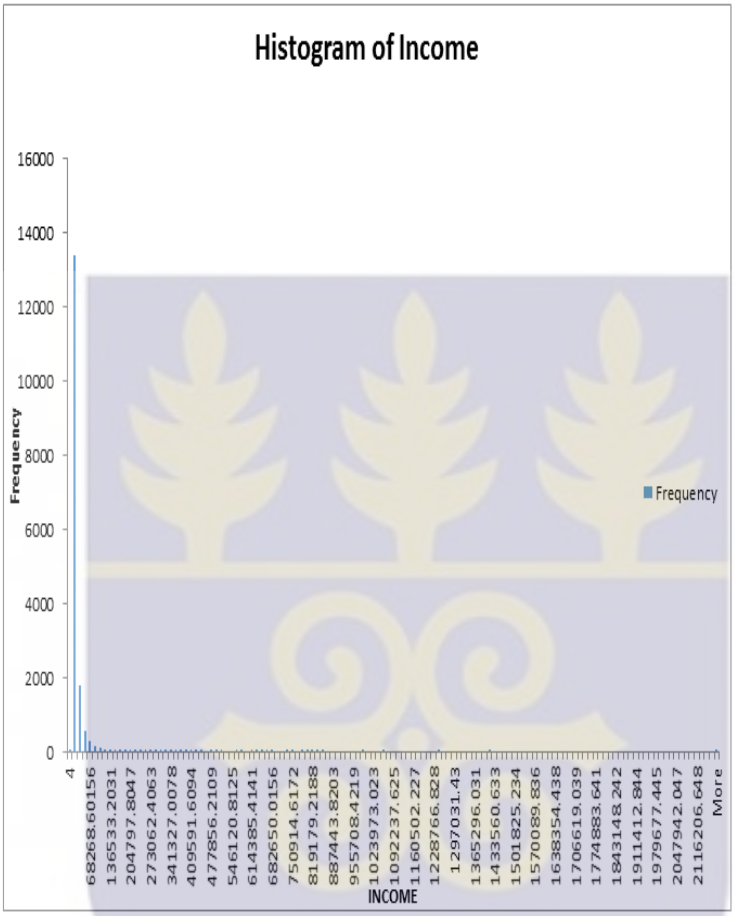


Figure 4.2: Histogram of Incomes

A Q-Q plot depicting the plotted quantile of the empirical distribution of income against the theoretical distribution which is used in assessing goodness of fit to the data set is illustrated in Figure 4.3. The curve represents the empirical distribution plot of income whereas the straight line depicts the plot of the quantiles of the lognormal distribution. It is observed that the empirical distribution plot of income line along the straight line which indicates that the income distribution is significantly the same as the lognormal.

University of Ghana http://ugspace.ug.edu.gh

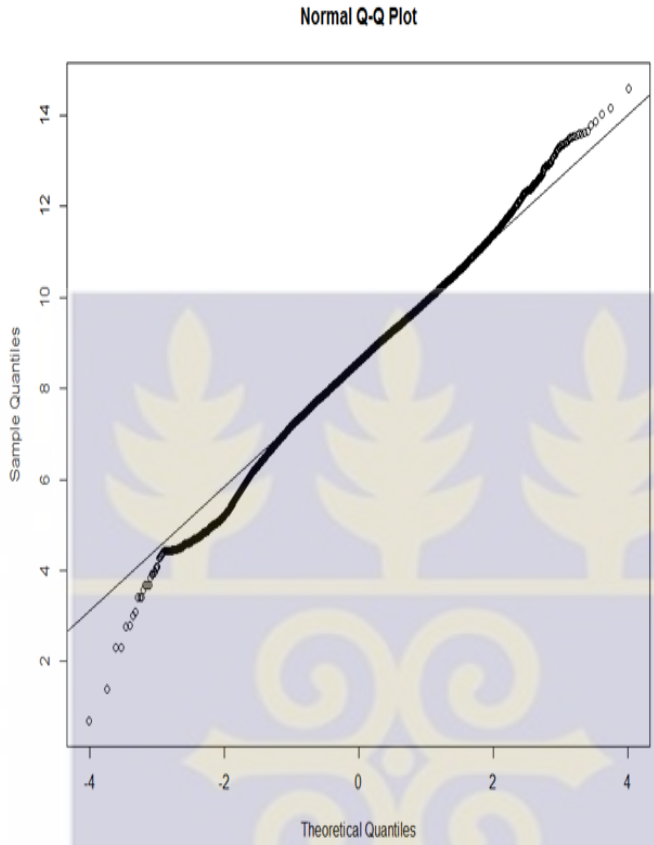


Figure 4.3: Lognormal Q-Q plot

The result from the Goodness of fit test with a value of 0.157 clearly support the conclusion that the distribution of the income is lognormal.Using a significance level of 0.05 , the Kolmogorov-Smirnov (KS) test produced p-values of 0.42 which is greater than 0.05 which support the conclusion that the distribution of the income is lognormal.

Consequently, the total income (x) follows the lognormal distribution with parameters . The density function of the income which follows the lognormal distribution is given by;

University of Ghana http://ugspace.ug.edu.gh

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(Inx-\mu)^2}{2\sigma^2}}, x > 0, \sigma > 0, \mu > 0 \tag{4.1}$$

The maximum likelihood estimation is used to estimate the parameters of the lognormal distribution given $\hat{\mu} = 8.5311$ and $\hat{\sigma} = 1.45812$

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

$$F(x) = \Phi\left(\frac{Inx - 8.5311}{1.45812}\right) \tag{4.2}$$

Given $\pi(x)$ to be the proportion of the units that receive income up to x and $\eta(x)$ to represent the proportion of total income received by the same units whose income are less than or equal to x. The Lorenz curve of income showing the relationship between $\eta(x)$ and $\pi(x)$ using the income data from GLSS6 is shown in Figure 4.3 below, where $\eta(x) = mcdf$

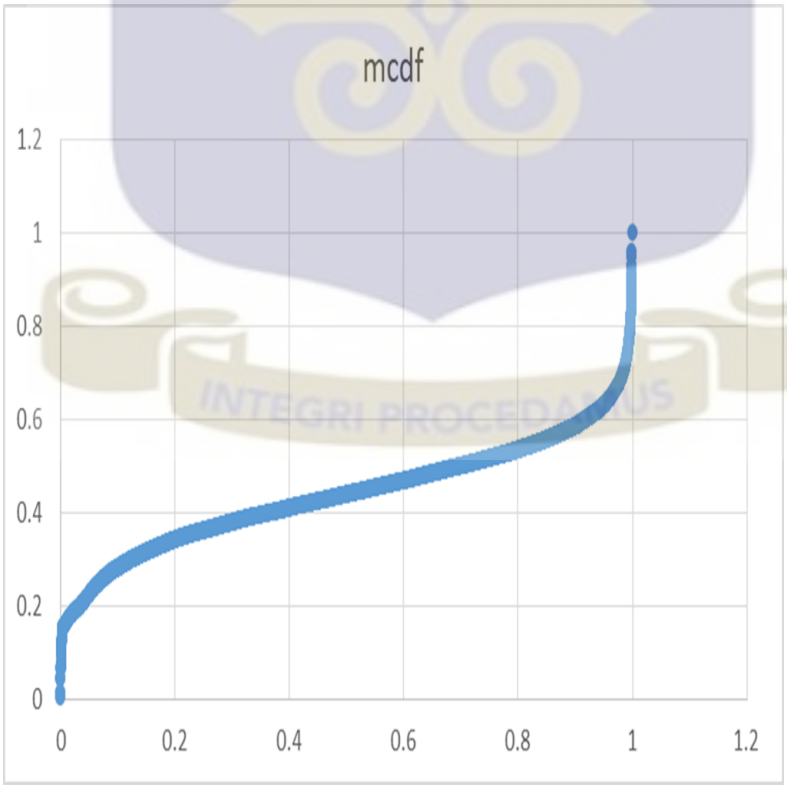


Figure 4.4: The Lorenz Curve of income

University of Ghana http://ugspace.ug.edu.gh

Rasche et.al.(1980) proposed the Lorenz curve estimation for nonlinear income distributions as;

$$\eta(x) = 1 - (1 - \pi)^{0.347}$$

The coefficient of determination of regressing $\eta(x)$ on $\pi(x)$ using Rasche et. al.(1980) method is 0.89. This indicate that eighty nine percent of the variance in the response variable can be explained by the explanatory variables. The Rasche et. al.(1980) function obtained is integrated in order to yield approximate area under the Lorenz curve. The Lorenz curve estimate is given by;

$$L(\pi) = \int_0^1 1 - (1 - \pi)^{0.347} d\pi$$

Let $x = 1 - \pi$. when $\pi = 0$, $x = 1$ and when $\pi = 1$, $x = 0$. Then

$$L(x) = -\int_1^0 1 - x^{0.347} dx$$

$$= -\left[x - \frac{x^{1.347}}{1.347}\right]_1^0$$

$$= \left[1 - \frac{1}{1.347}\right]$$

$$= 0.2575$$

Hence the Gini coefficient is estimate as

$$\text{Gini} = 1 - 2(0.2575)$$

$$Gini = 0.485$$

The estimate of the Gini coefficient of income applied to the Lorenz curve for the lognormal distribution by using the Rasche et.al.(1980) function on the

University of Ghana http://ugspace.ug.edu.gh

interval (0, 1) is given by 0.485.

Also,the coefficient of determination of regressing $\eta(x)$ on $\pi(x)$ using the polynomial function with degree 5 is 0.98. This indicates that ninety eight percent of the variance in the response variable can be explained by the explanatory variables.

The polynomial function with degree 5 by regressing η on π using the income data from GLSS 6 is given by;

$$\eta(\pi) = 0.987\pi^5 - 5.734\pi^4 + 4.993\pi^3 - 6.355\pi^2 + 4.225\pi \tag{4.3}$$

The polynomial obtained is integrated in order to yield approximate area under the Lorenz curve. The Lorenz curve estimate is given by;

$$L(\pi) = \int_0^1 0.987\pi^5 - 5.734\pi^4 + 4.993\pi^3 - 6.355\pi^2 + 4.225\pi d\pi \tag{4.4}$$

$$= \left[\frac{0.987}{6}\pi^6 - \frac{5.734}{5}\pi^5 + \frac{4.993}{4}\pi^4 - \frac{6.355}{3}\pi^3 + \frac{4.225}{2}\pi^2\right]$$

$$= 0.260117$$

Hence the Gini coefficient of income is estimated as;

$$= 1 - 2(0.260117)$$

$$= 0.480$$

The estimate of the Gini coefficient of gross income applied to the Lorenz curve for the lognormal distribution using the proposed integration method on the interval (0, 1) is given by 0.48. Slesnick (1994) confirms that income inequality measures inequality based on current gross income. The estimates of the Gini

University of Ghana http://ugspace.ug.edu.gh

coefficient of the gross income of the 10 administrative regions of Ghana,Rural and Urban areas and Male and Female household heads are shown in table 4.2 below. The Gini coefficients of the gross income for the regions are between 0.42 and 0.57. The region with the highest gross income disparity is the Upper west region and Greater Accra region has the lowest income disparity.Also,the Gini coefficient of income of Rural and Urban areas are 0.44 and 0.49 respectively which shows that the Urban area has a high level of income inequality as compared to the Rural area. Furthermore, the Gini coefficient of gross income of Male Household Heads and Female Household Heads are 0.51 and 0.46 respectively which also shows that the Male Household Heads have a high level of gross income inequality as compared to the Female Household Heads

Table 4.2: Estimates of the Area under the curve and Gini Coefficient of Income in the 10 administrative regions of Ghana

Region	π^5	π^4	π^3	π^2	π	Area	Gini
Greater Accra	1.416	-7.425	14.684	-9.208	0.34	0.29	0.42
Volta	0.654	-6.845	11.92	-5.295	0.64	0.275	0.45
Eastern	0.924	4.915	-9.916	-5.472	6.862	0.265	0.47
Central	1.914	-7.6	11.94	-5.628	0.744	0.28	0.44
Brong Ahafo	0.87	4.195	9.66	-7.893	-1.016	0.26	0.48
Ashanti	2.448	-6.625	7.444	-4.779	1.868	0.285	0.43
Upper East	1.284	-4.92	4.94	-1.686	0.634	0.22	0.54
Upper West	1.758	-8.42	-8.068	8.286	1.722	0.215	0.55
Northern	-5.904	14.61	-19.916	13.581	-2.492	0.24	0.51
Western	-4.377	12.381	-17.695	9.437	-0.397	0.27	0.46
Urban	6.745	11.657	-12.765	3.890	-4.965	0.25	0.5
Rural	-4.2575	-12.601	2.575	-5.753	7.439	0.27	0.46
Male Heads	9.2534	0.238	-5.871	-9.357	11.862	0.255	0.49
Female Heads	0.251	-6.870	3.854	8.675	1.675	0.265	0.47
All Ghana	0.987	-5.734	4.993	-6.355	4.225	0.26	0.48

The chapter presented the findings and discussions from the analysis of the study. The income data was displayed and the graphical quantile – quantile

University of Ghana http://ugspace.ug.edu.gh

plots of the logarithm income illustrating the distribution of the income was also presented. The one sample Kolmogorov-Smirnov test was also performed to demonstrate fitness to the lognormal distribution. The analysis established that the income distribution follows the lognormal distribution with scale parameter 5.8311 and shape parameter equals 1.456. Also the estimate of the Gini coefficient of income applied to the Lorenz curve for the lognormal distribution using the proposed integration method on the interval (0, 1) is given by 0.48.

The estimate of the Gini coefficient of gross income applied to the Lorenz curve for the lognormal distribution using the proposed integration method on the interval (0, 1) is given by 0.485. The Gini coefficients of the gross income for the regions are between 0.42 and 0.57, for Rural and Urban areas are 0.44 and 0.49 respectively and for Male Household Heads and Female Household Heads are 0.51 and 0.46 respectively. The next chapter presents summary, conclusion and recommendations of the research and also further study areas.

INTEGRI PROCEDAMUS

University of Ghana http://ugspace.ug.edu.gh

Chapter 5

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter presents a summary of the findings from the study and conclusions drawn from those findings. Some recommendations are also made including some directions for further study with possible areas for future research.

5.2 Summary of Findings

The main purpose of this research was to propose a better numerical integration method to estimate the area under the curve and Gini coefficient applied to the lognormal distribution. The study considered the income data from the Ghana Living Standard Survey, 2013. A preliminary analysis of data in the form of exploratory techniques revealed that the data is heavy tailed and hence

University of Ghana http://ugspace.ug.edu.gh

follows the lognormal distribution. Also, the Kolmogorov-Smirnov goodness of fit test accepted that the data is lognormal distributed.The study employed income data from the Ghana Living Standard Survey in the year 2013. The total gross income (x) of GH¢244,759,213.20 from a total family size of 76,269 was derived from the 2013 Ghana Living Standard Survey (GLSS).

Findings from the study discovered that the income distribution from the Ghana Living Standard Survey is leptokurtic, heavy tailed and asymmetrically distributed and can be modeled with a lognormal distribution. Hence the income data follows the lognormal distribution with $\mu = 8.5311$ and $\sigma = 1.45812$. The study shows that the share of the population living in poverty was reduced from 51.7 percent in 1991/92 to 39.5 percent in 1998/99 and 31.9 percent in 2005/2006 to 24.2 in 2012/2013. The numerical integration was used to calculate $\pi(x)$ which is the proportion of the units that receive income up to x and $\eta(x)$ which is the proportion of total income received by the same units whose income are less than or equal to$x$. The coefficient of determination of regressing $\eta(x)$ on $\pi(x)$ using polynomial of degree five and Rasche, Gaffney et. al.(1980) is 0.98 and 0.89 respectively.

The various numerical integration techniques such as the trapezium rule, Simpson's 1/3 rule, Simpson's 3/8 rule, Boole's rule and Weddle's rule were compared with the proposed numerical integration method in estimating the Lorenz curve and Gini coefficient. Estimates of the Lorenz curve using the polynomial function of degree 5 and Rasche et.al.(1980) are 0.260117 and 0.2575 respectively. Also,the estimate of the Gini coefficient of income applied to the Lorenz curve by using the polynomial function of degree 5 and the Rasche et. al.(1980) function on the interval (0, 1) is given by 0.4798 and 0.485 respectively. Comparing Rasche et.al.(1980) function to the Polynomial function of degree 5, the Polynomial function of degree 5 gives a better estimate since its

University of Ghana http://ugspace.ug.edu.gh

coefficient of determination gives a better explanation of $\pi(x)$ by $\eta(x)$.

5.3 Conclusion

The study propose a numerical integration method that provides a better estimate as compared to the Newton-Cotes methods of integration, and apply it to the stochastic estimation of the Gini coefficient. Also, comparing Rasche and Gaffney function to the Polynomial function of degree 5 to estimate the Lorenz curve, the Polynomial function of degree 5 gives a better estimate of the Lorenz curve since its coefficient of determination gives a better explanation of how the data fit the model. The Proposed numerical integration method is used to estimate the Lorenz curve and Gini coefficient of income.

Comparing the Proposed Numerical Integration Method to the various numerical integration formulas from Table 2 above using the errors, when $k = 1$, the area under the curve is 16.9531 which is the same as the area under the curve using the Trapezium rule. Also when $k = 2$ and above, the estimates of the area under the curve are better than the Trapezium rule. That is, the Trapezium rule yields the highest error of 0.5016 as compared to the exact integration. Furthermore, when $k = 6$ and above, the estimates of the area under the curve are better than the Simpson $1/3$and $3/8$ rule. When $k = 12$ and above, the estimates of the area under the curve are better than Boole's rule and when $k = 15$ and above, the estimates of the area under the curve are better than Weddle's rule. Finally, when $k = 20$, the area under the curve is 16.4539 which is closer to the exact value of the area under the curve, which is 16.4509 yielding 0.024 errors as compared to the other numerical integration methods. The proposed method with smaller divisions of the interval gives a better estimate with lesser errors as compared to the Trapezium rule, Simpson

University of Ghana http://ugspace.ug.edu.gh

1/3 and 3/8 rule, Boole's and Weddle's rule.

The Gini coefficient using the proposed numerical integration method with k=3 was 0.48 which shows that there is disparity in income in Ghana. The study therefore concludes that the proposed method is better than the Newton-Cotes methods of integration and it was found that the income data from the Ghana Living Standard Survey 6 can be modeled using the lognormal distribution.

5.4 Recommendations

It has been established in the study that the proposed numerical integration method in estimating the area under a curve proved to be more efficient compared to the popular or traditional methods such as the Trapezium rule, Simpson 1/3 rule, Simpson 3/8 rule, Boole's rule and Weddle's rule for integrating functions. Also, the proposed numerical integration method can be applied to an empirical data which follows a probability distribution function for estimating the Gini coefficient of income. It is therefore recommended to investors, policy makers, financial analysts and stakeholders to make use of the proposed numerical integration method when computing the inequality of income of a country. It is further suggested that statisticians or mathematicians should use the proposed numerical integration method when computing functions that cannot be easily integrated.

5.5 Further Studies

Directions for future studies include:

1. Research using the proposed numerical integration method to estimate

University of Ghana http://ugspace.ug.edu.gh

the Gini coefficient applied on different probability distributions such as Pareto distribution, exponential, beta etc

2. An application of the proposed numerical integration method to estimate the Gini coefficient in other areas such as education, agriculture, biodiversity, stock exchange in Ghana.

3. An investigation into the effectiveness of the Bayesian approach in estimating the Gini coefficient of income and comparing it with estimates obtained from using the proposed numerical integration method.

4. An application of the proposed numerical integration method to estimate the tails of the Lorenz curve.

University of Ghana http://ugspace.ug.edu.gh

Bibliography

[1] Aitchison, J., & Brown, J. (1957). *The Lognormal Distribution*. Cambridge, UK: Cambridge University Press.

[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] Anand,S. (1983). *Inequality and Poverty in Malysia : Measurement and Decomposition*. New York: Oxford University Press.

[4] Anand, S. (1997). *The measurement of income inequality*. In S. Subramanian, Measurement of inequality and poverty (pp. 81-105). Oxford University Press.

[5] Anand, S., & Nanthikesan, S. (2000). *The evolution of inequality in length of life : France in* 1806 – 1997. Working Papers Series, 11(3), Harvard Center for population and Development Studies, Cambridge, MA, February.

[6] Anand, S., Diderichsen, F., Evans, T., Shkolnikov, V., & Wirth, M. (2001). *Measuring disparities in health : methods and indicators*. In: Challenging Inequalities in Health: From Ethics to Action. New York: Oxford University Press: 49-67.

University of Ghana http://ugspace.ug.edu.gh

[7] Atkinson, A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory* 2, 244-263.

[8] Atkinson, A. B. (1970).*Poverty in Britain and the Reform of Social Security*. Cambridge University Press.

[9] Atkinson, A. B. (1980).*Wealth, Income and Inequality*, 2nd edn. Oxford: Oxford University Press.

[10] Atkinson, A. B. (1987). O n the measurement of Poverty. *Econometrica*, 55, 749-764.

[11] Balintfy, L., & Goodman, S. (1973). Socio-Economic Factors in Income Inequality: A Lognormal Hypothesis. *Zeitschrift fiir Nationalokonomie*, 389-402.

[12] Bassman, R., Hayes, K., Johnson, J., & Slottje, D. (1988). A General Functional Form for Approximating the Lorenz Curve. *Journal of Econometrics* 43, 77-90.

[13] Becker, T., & Weispfenning, V. (1993). *A computational approach to commutative algebra*. New York: Springer-Verlag.

[14] Berndt, D. J., Fisher, J. W., Rajendrababu, R. V., & Studnick, J. (2003). Measuring Health Care and Inequities Using the Gini Index. *36th Annual Hawaii International Conference on System Sciences*, (pp. 159-163).

[15] Berrebi, M., & Silber, J. (1985). "Income Inequality Indices and deprava- tion: A Generalization", *Quarterly Journal of Economics*, 99, 807-810.

[16] Berrut, J. P., & Trefethen, L. N. (2004). Barycentric Lagrange Interpolation. *SIAM Review*, 46(3), 501-517.

University of Ghana http://ugspace.ug.edu.gh

[17] Blaylock, J., & Smallwood, D. (1982). Analysis of Income and Food Expenditure Distribution. *Review of Economics and Statistics* 64, 104-109.

[18] Chakravarty, S. (1988). *Extended Gini Indices of Inequality*. International Economic Review, vol.29, Issues 1, 147- 156.

[19] Cheng, B. (2005). Lorenz curve and Gini coefficient under condition of log-normal distribution.*Journal of Quantitative and Technical Economics* 2, 127-135.

[20] Chotikapanich, D., & Griffiths, W. E. (2002). Estimating Lorenz curve using a Dirchlet distribution.*Journal of Business and Economic Statistics* 20, 290-295.

[21] Conover, W. J. (1999). *Practical Nonparametric Statistical*. New York: John Wiley & Sons, Inc.

[22] Cowell, F. A. (1977). *Measuring Inequality*. Oxford: Phillip Allen.

[23] Cramer, F. (1978). A function for the size distribution of income. *Econometrica* 46, 459-460.

[24] Dalton, H. (1920). Measurement of the Inequality of Income. The Economic Journal, 30, 348-361.

[25] Damgaard, C., & Weiner, J. (2000). Describing inequality in plant size or fecundity .*Ecology* 81(4), 1139-1142.

[26] Donaldson, D., & Weymark, J. (1980). A single Parameter Generalization of the Gini Indices of inequality. *Journal of Economic Theory* 22: 67-86.

[27] Dorfman, R. (1979). A Formula for Gini Coefficient.*Review of Economics and statistics* 61,146-149.

University of Ghana http://ugspace.ug.edu.gh

[28] Edozien, E. C. (1991). *Inequality and Poverty Issue in Concept and Thorny*. The Nigeria Economic Society.

[29] Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). Modelling Extreme Events for Insurance and Finance. Heidelberg: Springer-Verlag.

[30] Evans, M., & Swartz, T. (1995). Methods for aproximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science* 10, 254-272.

[31] Fellman, J. (2012). Estimation of Gini coefficients using Lorenz Curves. *Journal of Statistical and Econometric Methods*, vol.1 no. 2, 31-38.

[32] Foster, J. E., & Sen, A. (1997). *On economic inequality after a quarter century*. Oxford,UK: Clarendon Press.

[33] Foster, J., Greer, J., & Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica* 52(3), 761-766.

[34] Garner, T. (1993). Consumer Expenditures and Inequality:An Ananlysis Based on Decomposition of the Gini Coefficient. *The Review of Economics and Statistics*, 134-138.

[35] Gastwirth, L. J. (1972). The Estimation of the Lorenz Curve and Gini Index. *The Review of Economics and Statistics*, *LIV*, *No*.3, 306–316.

[36] Gibrat , R. (1931). *Les inegalites economiques*. Sirey: Paris.

[37] Gini, C. (1912). Measurement of Inequality of Incomes. *The Economic Journal* 31, 124-126.

[38] Giorgi , G. M., & Pallini, A. (1987). A general method for the lower and upper distribution free bounds on Gini's concentration ratio from grouped data.*Statistica*, 47, 171-184.

University of Ghana http://ugspace.ug.edu.gh

[39] Glewwe, P. (1985). The distribution of Income in Sri Lanka in 1969-70 & 1980-81: *A Decomposition Analysis*" *Department of Agricultural Economic and Rural Sociology*, Pennsylvania State University.

[40] Gnanadesikan, R., & Wilk, M. B. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55, 1-17.

[41] Golden, J. (2008). A simple geometric approach to approximating the Gini coefficient. *Journal of Economic Education*, 39(1), 68-77.

[42] Goldie, C. M. (1977). Converegence theorems for empirical Lorenz curves and their inverses, *Advances in Applied Probability* 9. 765-791.

[43] Goodwin, D. G., & Vaupel, J. W. (1985). *Curves and have – statistics forecological analysis of diversity. Part III : Comparison of measures of diversity.* Laxenburg,Austria: International Institute for Applied System Analysis.

[44] Gordon, K. S. (1998). Numerical Integration.*Encyclopedia of Biostatistics*: ISBN: 0471975761.

[45] Graham. (1995). *Poverty and Health Risks*; Issues and Concept.

[46] Gupta, M. R. (1984). Functional Form for Estimating the Lorenz Curve. *Econometrica*, 52, 1313-1314.

[47] Heshmati, A. (2004). The World Distribution of Income and Income Inequality.*IZA Discussion Paper No.*1267.

[48] Holzmann, H., Vollmer, S., & Weisbrod, J. (2003). Income distribution dynamics and pro-poor growth in the world from 1970 to 2003. *Ibero America Institute for Economic Research (IAI) Discussion papers*, 161.

University of Ghana http://ugspace.ug.edu.gh

[49] Husted, T. (1991). Changes in State Income Inequality from 1981 to 1987. *TheReview of* 135 *Regional Studies*, *Number* 3, 248-260.

[50] Iyengar, N. S. (1960). On a Method of Computing Engel Elasticities from Concentration Curves. *Econometrica*, 28(4), 882-891.

[51] Jain, S. B. (1975). *Size Distribution of Income.A computation of Data*. Washington: The World Bank.

[52] Kakwani, N. (1978). A New Method of Estimating Engel Elasticities. *Econometrica*, 8, 103-110.

[53] Kakwani, N. (1980). On a Class of Poverty Measures.*Econometrica*, Vol.8, No.2, 437-446.

[54] Kakwani, N., & Podder, N. (1973). On the Estimation of Lorenz Curves from Grouped Observations. *International Economic Reviews*, Vol.14, no.8, 278-292.

[55] Kalecki, M. (1945). On the Gibrat Distribution. *Econometrica*, 13, 161-170.

[56] Kendall, M. G., & Stuart, A. (1958). *The AdvancedTheory of Statistics*, Vol 1, Distribution Theory. New York: Hafner Publishing Company.

[57] Kratz, M. F., & Resnick, S. I. (1995). *The QQ − Estimator and HeavyTails*. New York:Cornell University: Discussion Paper,School of ORIE.

[58] Labor, U. D. (1978).*Consumer Expenditure Survey* : *Interview Survey*. Bureau of Labor Statistics.

[59] Lerman, R. I., & Yitzhaki, S. (1984). A Note on the Calculation and Interpretation of the Gini Index. *Economics Letters*, 15, 363-368.

University of Ghana http://ugspace.ug.edu.gh

[60] Levy, F. (1987). Changes in the distribution of American Family Income,1947 to 1984. *Science* 236, 923-927.

[61] Lopez, J. H., & Serven, L. (2006). A Normal Relationship? Poverty,Growth and Inequality. *World Bank Policy Research Working Papers*, 3814.

[62] Lorenz, M. O. (1905). Methods of Measuring the Concentration of Wealth. *Journal of American Statistical Association*, 9, 209-219.

[63] Lorenzen, G. (1980). Was ist ein, echtes"Konzentrations ma B"? *Allgemeines Statistiches Archiv*, 4, 390-400.

[64] Mandelbrot, B. (1960). The Pareto-Levy law and the distribution of income.*Econometrica* Rev 1, 79-106.

[65] McDonald, J. B. (1984). Some Generalized Functions for the Size Distribution of Income. *Econometrica Vol*.52, *No*.3, 647-665.

[66] McDonald, J. M., & Xu, Y. J. (1995). A generalization of the beta distribution with applications.*Journal of Econometrics* 66, 133-152.

[67] Mehran, F. (1978). Bounds on the Gini Index based on observed points of the Lorenz curve. *Journal of America Statistics Association*, 70, 64-66.

[68] Morgan, J. (1962). The anatomy of income distribution.*The Review of Economics and Statistics*, 44 , 270-283.

[69] Needleman, L. (1978). On the approximation of the Gini coefficient of concentration. *The Manchester School*, 105-122.

[70] Nolan, J. P. (2003). *Modelling financial data with Stable Distributions*. In S. Rachev, Handbook of Heavy Tailed Distributions in Finance (pp. 107-129).

University of Ghana http://ugspace.ug.edu.gh

[71] Norman, J., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distribution*, Vol 1, 2nd edn. New York: John & Sons.

[72] Pareto, V. (1897). *Cours d'Economie Politique*. Lausanne.

[73] Pestieau, P., & Possen, U. (1979). A model of wealth distribution. *Econometrica* 47, 761-772.

[74] Pinkovskiv, M. (2008). *Testing Lognormal Mixtures against the Generalized Beta Distribution as a Parametric Model for the Distributions of Income and Expenditure*. Columbia mimeo.

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

[76] Rowntree, B. F. (1901). *Poverty A Study of Town Life*. London: MacMillan.

[77] Rutherford, R. S. (1955). Income Distributions: A New Model. *Econmetrica*, 23, 277-294.

[78] Ryu, H. K., & D.J.Slottje. (1996). Two flexible functional form approaches for approximating the Lorenz curve. *Journal of Econometrics* 72, 251-274.

[79] Ryu, H., & Slottje, D. (1999). *Parametric Approximations of the Lorenz Curve*. In J. Silber, Handbook of Income Inequality Measurement (pp. 291-314). Springer.

[80] Sala-i-Martin, X. (2006). *The World Distribution of Income* (Estimated from Individual Country Distributions). NBER working Paper, 8933.

University of Ghana http://ugspace.ug.edu.gh

[81] Sargan, D. (1957). The Distribution of Wealth. *Econometrica* 25, 568-590.

[82] Sen, A. K. (1973). *On Economic Inequality*. Oxford: Claredon Press.

[83] Sen, A. (1984). *Resources, Values and Development*. Oxford: Basil Blackwell.

[84] Shirras, G. F. (1935). The Pareto Law and the distribution of Income. *EconomicJournal* 45, 663.

[85] Shorrocks, A. F. (1973). Ranking Income Distributions. *Econometrica*, 50, 3-17.

[86] Silber, J. (1989). Factors Componen ts,Population Subgroups and the Computation of the Gini Index of Inequality. *The Review of Economics and Statistics*, *LXXI*107-115.

[87] Slottje, D. J. (1989). *The Structure of Earnings and the Measurement of Income Inequality in the US*. Amsterdam,North Holland.

[88] Souma, W. (2000). Universal Structure of the Personal Income Distribution. *arXiv* : *condmat*/0011373*v*1.

[89] Steindl, J. (1965). *Random Process and the Growth of firms* : A study of the Pareto Law. New York: Hafner.

[90] Steyn, S. (1959). Spatial Decomposition of Inequality. *South American Journal of Economics*, 27(2), 149-156.

[91] Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North-Holland Pub.

University of Ghana http://ugspace.ug.edu.gh

[92] Wilmoth, J. R., & Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography* 36(4), 475-495.

[93] Xiao, D., Feng, X., & Shiqiang, Z. (2014). A Simple and Practical Method of Calculating the Gini Coefficient. *Progress in Applied Mathematics*,Vol 8, No.1, 29-33.

[94] Xu, K. (2004). *How has the literature on Gini's index evolved in the past 80 years*? Dalhousie: University Retrieved on June 1,2006.

[95] Xu, K., & Osberg, L.(2002). The Social Welfare Implications, Decomposability, and Geometry of the Sen Family of Poverty Indices. *Canadian Journal of Economics*, 35, 138-152.

[96] Yitzhaki, S. (1990). *On the Progressivity of Commodity Taxation*. Jerusalem, Israel: Hebrew University.

[97] Yitzhaki, S., & Thrisk. (1990). Welfare Dominance and Design of Excise Taxation in the Cote D'Ivoire. *Journal of Development Econmics* 33, 1-18.

[98] Yitzhaki, S., & Schechtman, E. (2005). The Properties of the Extended Gini Measures of Variability and Inequality. *Social Science Research Network*.