STATISTICAL MEASURES

OF

INCOME DISTRIBUTION WITH REFERENCE TO RURAL ETHIOPIA

A THESIS

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES

ADDIS ABABA UNIVERSITY

IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN STATISTICS

BY

MILLION DENBEL

JUNE, 1989

TABLE OF CONTENTS

PAGE

PAGE

Appendix:

Acknowledgements

I am heavly indebted to my advisor, Asmerom Kidane, Associate Professor, Head Department of Statistics, for his valuable advice and for his unforgetable supply of reference materials which are essential to this study.

I would like to thank my employer, Agricultural and Industrial Development Bank of Ethiopia who granted me sponsorship for MSC Programme and in particular my thanks go to Ato Mesfin Gebeyehu, Head Research and Planning Services, and Ato Mulugeta Tesfaye, Head Statistics Division, for their unreserved encouragement in the course of this work.

I am also indebted to Ato Genene Zewge of ECA for his supply of relevant reference material.

Finally, I should also extend my thanks to Wt. Sirgute G/Yesus and W/ro. Senait Belete for typing the manuscript.

ABSTRACT

Statistical methods of measuring the distribution of income have been investigated as far as computing facility is available. The data set utilized in this study is the sub-sample income data of the agricultural sedentary population of Ethiopia excluding Tigrai and Eritrea. The data were collected by the Central Statistics Authority and covers a period of one year, from May 1981 to April 1982.

In particular, among others, emprical results for four inequality measures, namely, Gini cofficient, Theil's entropy index, Relative mean deviation and logarthmic variance have been obtained. moreover the liability of the lognormal and the pareto distributions as a mathematical discription of the income data under consideration have been checked. Based on the computed results, some conclusions are made.

1. INTRODUCTION

1.1 Objective of the Study

The most common indicator of economic development is the per capita income. However, such a measures is believed to be less satisfactory in a number of ways one of which is its inability to take into account the distributional aspect of income among a society.

In the last several decade studies on the relationship of economic growth and income distribution at different stages of economic development have been undertaken, Arthur(1985). One of the best known is the works of Simon Kuznets, after with certain extensive supports, has postulated an inverted u-shaped hypothesis exists between economic growth and income distribution. That is, the distribution of income to be more unequal in the early stage of economic growth, increases as the economy progresses, and becoming low at a higher stages.

There are a number of factors affecting the distribution of income, Sundrum, (1974). But, the study of factors affecting distribution of income and in addition, its changes overtime is mostly handicapped by severe lack of data. This problem may be highly aggravated in developing countries than developed countries, and hence many objectives regarding the study of income distribution are of limited scope.

Although various studies have been carried in different parts of the world, both developing and developed countries, here in Ethiopia, it may be due to scarcity of well compiled data, very little of such type of study is conducted.

When dealing with measurement issues, the study of income distribution among a society rely on many concepts from statistics. Thus, the objective of this paper is:

i) To apply various statistical measures of income distribution to the existing income data set of Ethiopia

ii) To identify the relevance of the lognormal and the pareto distribution as a mathematical description of the income data under consideration.

Chapter two is devoted to description of methods that could possibly be applied in the study of income distrbution. Source and an overview of the data are given in Chapter three. Chapter four deals with Results and Chapter five is confined to Discussion and Conclusion.

2. METHODS

There are a number of statistical measures of income inequality each capturing a different characterstic of the social state and possessing desirable properties as a yardstic of inequality.

The inequality measures, in general, are classified as positive measures and normative measures. The main difference between these two classes of measures is that, the latter make explicit use of social welfare function while the former do not. Under social welfare function, all the possible states of society are ranked in order of society's preference, Cowell (1977). However, in some ways the positive measures may be viewed as normative measures under certain specific assumptions of the social welfare function.

The positive measures include such as Range, Mean deviation, variance, coefficient of variation, logarthmic variance, variance of logarthms, Gini Coefficient. Thiel's entropy index etc, whereas the normative measures include Dalton's index and Aitinkson's index.

Therefore, in this Chapter, properties of the measures in capturing inequality of income and their estimation are considered. Moreover, two theoretical distributions, the lognormal and pareto distributions, liability as a mathe-

matical descriptions of income data, estimation of their parameters, how to test their suitability and application of them in relation to measuring inequality are discussed in the last Section of this Chapter. However, implementation of these measures to our income data depends on the nature of the data, importancy of the measures and availability of computer facility.

2.1. Some Simple Measures

The simplest measure of inequality is the Range which is based on comparing the extreme values of the distribution. The problem with the Range is that by concentrating on extreme values only it misses the important features of the distribution. It might be satisfactory in a closed society where everyone's income is fairly known. Another way of looking at the distribution is by comparing each income with the mean income level, that is, by Relative mean deviation. The disadvantage of this measure is that it is not sensitive to transfers of income from poor to rich as long as both lie in the same side of the mean income level since the transfer would add one gap and reduce another gap by exactly the same amount.

Variance, the common statistical measure of variation, is also used as inequality measure. Variance as inequality measure has an attractive property when dealing with transfer

of income from the poorer to a richer person. This transfer definitely alter the variance. This property has to be the minimum property that has to be possed by inequality measures, H. Dalton, 1920.

However, even if the variance satisfies the condition pointed above, it is unsatisfactory measure of inequality because if everyone's income is doubled; the shape of the distribution will remain unchanged. Also since, the variance depends on the unit of observation one distribution may show much greater relative variation than another and still end up having a lower variance. This difficulty is overcomed by standardizing the variance i.e. by taking the Coefficient of variation. The coefficient of variation is liable for capturing income transfer. But, an equal amount of transfer at any income level reduces the measured inequality by the same amount. Thus, it may reasonably argued that the coefficient of variation is good at capturing inequality among high income groups.

Contrary to capturing inequality among high income groups, there appears to be good reason to suggest inequality measure that are more effective in reducing inequality for income transfers when effected in low income brackets. The logarthmic variance and variance of logarithms of income appear to overcome this problem. The difference between these two measures lies in that the former is defined relative to the logarithm of income while the latter is

defined relative to the mean of logarithm of income.

Assuming there are n people and observation on income is on individual basis then the logarthmic variance, v, and variance of logarithms of income v_1, are defined as

$$V = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{X_i}{\bar{X}}\right)\right]^2 \qquad (2.1.1)$$

and

$$V_1 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{X_i}{\bar{X}^*}\right)\right]^2 \qquad (2.1.2)$$

where X_i is income of an individual
$\bar{X}$ is overall mean income
$\bar{X}^*$ is mean of logarithms of income.

These two measures attach greater importance to transfers in low income brackets. In other words, a transfer of income in low income brackets reduces v or v_1 more than a transfer in high income brackets do. One other advantage of taking logarithm, unlike the variance of actual values is that it eliminates the arbitrary of units and therefore of absolute levels, since a change of units, which takes the form of multiplication of the absolute values, comes out in a logarithm form as an addition of a constant and goes out when pair wise differences are taken.

However, it is possible for v or v_1 to rise even when there is a rich to poor transfer in high income bracket which is not a desirable property to be possesed by inequality measures.

2.2 Gini Coefficient

The most commonly used inequality measure is the Gini Coefficient attributed to Gini (1912), and much analysed by Ricci (1916) and later by Dalton (1920), Ynetema (1953), Aitinkson (1970), Newbery (1970), Sheshiniski (1972) and others. One way of looking it is interms of the Lorenz Curve; whereby the percentage of population arranged from the poorest to the richest are represented on the horizontal axis and the percentage of income enjoyed by the population proportion is shown on the vertical axis. If the given percentage of population receive the same proportion of income, then there would be a line of equality specified by a 45 degree. Hence the Gini Coefficient is equal to the ratio of the area under the Lorenz Curve to the area under the line of equality which is half.

The Gini Coefficient can be obtained from the following formula when observations on income are made/reported for n individuals

$$G = \frac{1}{n^2\bar{X}} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_i - X_j| \qquad (2.2.1)$$

where $\bar{X}$ is overall mean income

X_i is income of individual i

X_j is income of individual j

In case of grouped data, the Gini Coefficient lower and upper bounds are given by:

lower bound: $$G_L = 2 \sum_{i=1}^{k} \sum_{j \geq i}^{k} \frac{n_i n_j}{n^2 \bar{X}} |\mu_i - \mu_j| \qquad (2.2.2)$$

upper bound: $$G_u = G_L + \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{n_i^2}{n^2 \bar{X}} \lambda_i [\mu_i - \ell_i] \qquad (2.2.3)$$

where K is number of classes or groups

n_i and n_j are number of persons in group i and group j respectively

μ_i and μ_j are mean income of group i and group j respectively

$\bar{X}$ is overall mean income

ℓ_i is lower limit of group i

λ_i is proportion of the population in group i getting income ℓ_i.

Formulas (2.2.2) and (2.2.3) are derived by using the assumptions in Section 2.5 to be discussed later.

The Compromise value for the Gini which works for most theoretical distributions, is approximated by:

$$G = 2/3\ G_u + 1/3\ G_L \qquad (2.2.4)$$

The expression given above is found in Cowell,(1977)

There are many criticisms of the Gini coefficient. As Simon Kuznets argued the standard Gini coefficient is a summary that can conceal as much as it reveals and can obscure some of the major underlying factors that influence income distribution overtime. Also where relationships among the distribution of different economic variables are

of their probabilities and if we want to be able to add up the information values of the message regarding independent events we want h to have the property.

$h(p_i p_j) = h(p_i) + h(p_j)$, $i \neq j$ which can be satisfied if h is a function of the form h=-Log p.

Aggregating the information values of n events into a single number describes whether the entire system is more or less orderly. The aggregation is done by taking the weighted sum of the information values for the n possible events, where the weights being the respective probabilities. This average is known as the entropy of the system .i.e.

$$\text{Entropy} = \sum_{i=1}^{n} p_i h(p_i)$$

$$= -\sum_{i=1}^{n} p_i \log p_i \qquad (2.3.1)$$

Theils states that the above formula can provide a useful inequality measure.

Reinterpretation of the n possible events as n people in the population and the probability, p_i, as the share of an individual in the population, say s_i, allow us to utilize the entropy as a tool of inqquality measure. Moreover, if all people get an even share s_i equals 1/n. This value of s_i provide the maximum value of the entropy. Therefore, the Theils entropy index is given by the maximum value of the entropy minus the actual entropy that could be obtained from the data . Hence, for n

individuals, income data Theil's entropy index is given by:

$$T = \sum_{i=1}^{n} \frac{1}{n} h\left(\frac{1}{n}\right) - \sum_{i=1}^{n} S_i h(S_i)$$

$$= \sum_{i=1}^{n} S_i \left[h\left(\frac{1}{n}\right) - h(S_i)\right]$$

$$= \sum_{i=1}^{n} S_i \left[\log S_i - \log\left(\frac{1}{n}\right)\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[\frac{X_i}{\bar{X}} \cdot \log \frac{X_i}{\bar{y}}\right] \qquad (2.3.2)$$

where X_i is income of an individual i

$\bar{X}$ is overall mean income.

Transfer of income from rich to poor decreases T. The reduction in T depends on the ratio of the rich man income share to the poor man income share. Thus, if transfer of income at any income level is at same ratio the reduction in T is the same.

2.4 Social Welfare Function Based Inequality Measures

Another way of measuring inequality is in connection with the social welfare function. The two important classes of inequality measures derived based on the social welfare function are the Dalton(D_E) and Atinkson (A_E) inequality measures where E is the inequality aversion parameter which describes the strength of our yearning for equality vis-a-vis uniformly higher income for all Cowell,(1977).

Dalton views inequality as how far the actual average utility falls short of potential average social utility (If all incomes were distributed equally). Atinkson criticized the use of D_E on the ground that it is sensitive to the level from which the utility is meaured.

Given five properties of social welfare function*, that are:

1) Social welfare function (SWF) is individualistic and non-decreasing
2) SWF is symmetric, i.e. the value of SWF does not depend on the particular assignment of ~~l~~ables to members of the population
3) SWF is additive
4) SWF is strictly concave - the welfare weights decreases as x_i increases
5) SWF has constant elasticity, or constant relative inequality aversion if utility function $U(x_i)$ can be written as

$$U(x_i) = \frac{1}{1-E} x_i^{1-E}, \quad x_i \text{ is income of man i}$$

then we can obtain D_E and A_E from the following formulas

$$D_E = 1 - \frac{\sum_{i=1}^{n} x_i^{1-E}}{n\bar{x}^{1-E}} \qquad (2.4.1)$$

* Details are given in Cowell, (1977)

$$A_E = 1 - \frac{1}{\bar{X}} \left[\sum_{i=1}^{n} \frac{1}{n} X_i^{1-E} \right]^{\frac{1}{1-E}} \qquad (2.4.2)$$

where X_i is income of individual i

$\bar{X}$ is overall mean income

E is inequality aversion parameter

D_E given as (2.4.1) is only for $E \leq 1$. For $E > 1$ it is better to use $\frac{D_E}{D_{E-1}}$ which is bounded between 0 and 1. For $E=1$ it is recommended to replace X_i^{1-E} and $\bar{X}^{1-E}$ by $\log X_i$ and $\log \bar{Y}$, respectively, Cowell (1977).

2.5. Estimation of Inequality Measures from grouped income data

We have seen basic inequality measures and estimating them. But, most of the formulas are applicable when income data are given as individual records. In real world situation income data are available in grouped form. In such cases the formulas given in the preceeding Sections, except for the Gini coefficient, cannot be applied. Therefore the following approach may be utilized for grouped income data.

Suppose that for a particular population, the theoretical density function which gives the proportion

of population that has an income in infinitestimal interval x to x + dx is known. Again let us suppose that the desired inequality measure, or ordinally equivalent transformation of the desired inequality measure can be written in the form

$$J = \int_0^\infty h(X) f(X) dX \qquad (2.5.1)$$

where $h(X)$ is the desired inequality measure

$f(X)$ is the theoretical density function

For individual records of income data, for n people in the population, equation (2.5.1) reduces to its equivalent form as

$$J = \frac{1}{n} \sum_{i=1}^{n} h(Y_i) \qquad (2.5.2)$$

For grouped data, as given by equation (2.5.1), the value of J may be estimated by fitting a theoretical distribution that could possibly describe the income data. However, this approach may not be possible for reason that the theoretical distribution is a poor fit to the income data at hand. Therefore, no reliable estimates of inequality measure is obtained. If such a failarity is attained, overcoming this problem is to fit the pareto density function in each group, and J will have the form

$$J = \sum_{i=1}^{k} \int_{\ell_i}^{\ell_{i+1}} h(Y)g(Y)dY \qquad (2.5.1)$$

where k is number of classes/groups

ℓ_i is lower boundary of group i

ℓ_{i+1} is upper boundary of group i

Again in case of grouped income data it is possible to obtain the lower and upper bounds of inequality measures. These bound can also be used as a test criterion for suitability of theoretical distribution as a good fit to income data. This concept is to be discussed when we are dealing with theoretical distribution in the next Section.

The lower bound of an inequality measure of for grouped data is obtained by the following assumptions. If the class intervals and average income within class or group is given, we assume that everyone in each class gets the average income in that class i.e. we assume that there is no inequality within classes. If there are n_i people in class i, n people in the population, and these people are classified in k classes, then the lower bound, say J_L of an inequality measure is given by

$$J_L = \sum_{i=1}^{k} f_i h(\mu_i) \qquad (2.5.4)$$

where f_i is proportion of people in group i.

$h(\mu_i)$ is the desired inequality measure as a function of the average group income μ_i.

And to compute the upper bound, say J_u, we assume there is maximum inequality within each class, subject to the condition that the assumed average income within the class is equal to the observed average income of each class. So we assume that in class i everyone gets either the lowest or the highest income in that class. If we let the proportion λ_i of the people in class i stuck at the lower limit of class i, and $1-\lambda_i$ on the upper unit of class i, then,

$$J_u = \sum_{i=1}^{n} f_i [\lambda_i h(\ell_i) + (1-\lambda_i) h(\ell_{i+1})] \qquad (2.5.5)$$

where, λ_i, according to the assumption given above, can be shown to be

$$\lambda_i = \left(\frac{\ell_{i+1} - \mu_i}{\ell_{i+1} - \ell_i}\right)$$

and $h(\cdot)$ is the desired inequality measure defined on the upper and lower boundaries of each group.

2.6. Theoretical Distributions

There are two distributions that their application in measuring inequality of income is widespread. These are the lognormal and the pareto distributions. The lognormal distribution is said to be close approximation in low income range while pareto is said to be a good fit in the upper income range, Aitchison & Brown, (1963).

2.6.1 Lognormal distribution

Three types of lognormal distribution are identified; two parameter, three parameter and four parameter lognormal distributions. The distribution functions and their statistical properties are given in Aitchison and Brown (1963). However, in this paper the two parameter lognormal distribution, statistical properties, method of estimation of the parameters and test of lognormality is considered. Moreover, the reason behind why the lognormal distribution is considered as mathematical description of income data is discussed.

Let x be a positive variate defined in the interval $(0, \infty)$ such that $y = \text{Log } x$ is normally distributed with mean u and variance δ^2. Then x is said to be lognormally distributed with parameters μ and δ^2 and its distribution function is given by

$$\Lambda(X;\ \mu, \delta^2) = \int_0^X \frac{1}{X\sqrt{2\pi\delta}} \exp\{-\frac{1}{2\delta^2}(\log X - \mu)^2\}\ dX \qquad (2.6.1.1)$$

The distribution posseses moment of any order arising from the properties of moment generating function of the normal distribution. Hence, the j^{th} moment about the origin denoted by u_j is

$$\mu_j = \int_0^\infty X^j d\ (X)$$

$$= \int_{-\infty}^{\infty} e^{jy} dN(y)$$

$$= e^{ju+\frac{1}{2}j^2\delta^2} \qquad (2.6.1.2)$$

Therefore, the mean α and variance β^2 of the two parameter lognormal distribution is thus,

$$\alpha = e^{\mu+\frac{1}{2}\delta^2} \qquad (2.6.1.3)$$

$$\beta^2 = e^{2\mu+\delta^2}(e^{\delta^2}-1)$$

$$= \alpha^2\eta^2 \qquad (2.6.1.4)$$

where $\eta^2 = e^{\delta^2}-1$

The moment about the mean may readly found from the moment about the origin. In particular, the third and fourth moments about the mean, denoting by μ_3' and μ_4' respectively are

$$\mu_3' = \alpha^3(\eta^3+3\eta^4) \qquad (2.6.1.5)$$

$$\mu_4' = \alpha^4(\eta^{12}+3\eta^{10}+15\eta^8+16\eta^6+3\eta^4) \qquad (2.6.1.6)$$

In economic data skew frequency curves are the rule rather than exception. The lognormal distribution which is skewed has mainly, (Cowell, 1977), four reasons to be considered as a mathematical description of income data First, lognormal distribution has a lot of convenient properties such as its simple relation with the normal curve, symmetical Lorenz Curves, non-intersectingLorenz Curves, easy interpretation of parameters and reservation under log-linear transformation. Second, under certain kinds of random process the distribution of incomes eventually turns to be approximately lognormal. This idea here is that the change in peoples income can be linked to a systematic process where by, in each moment of time a person's income increases or decreases by a certain proportion, the exact proportionate increase or decrease being determined by chance. If the distribution of these proportionate increment or decrement follows the normal law, then the overall distribution of income approaches lognormality, provided enough time is allowed for the process to operate. Third. There is still some notion of individual utility or social welfare associated with logarithm of income; it would be nice to claim that although incomes do not follow the normal distribution, 'utility' or 'welfare' does. Fourth. the lognormal provides a reasonable sort of fit to many actual set of data.

likelihood, method of moments, method of quantiles and graphical methods.

Let $x_1, x_2, \ldots, x_n$ be sample observations of size n from $\Lambda(\mu, \delta^2)$. Then the j^{th} sample moment about the origin denoting by m_j and the respective sample moment about the mean denoting by m'_j are given by

$$m_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j \qquad (2.6.1.9)$$

$$m'_j = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i)^j \qquad (2.6.1.10)$$

we can also write

$$\bar{x} = m_i$$

$$v_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{n}{n-1} m'_2 \qquad (2.6.1.11)$$

and if we take log x = y, then

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad (2.6.1.12)$$

$$v_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad (2.6.1.13)$$

The likelihood function of the two parameter lognormal distribution is:

$$L(\mu, \delta^2; x) = \frac{1}{\delta^n (2\pi)^{\frac{n}{2}} \prod_{i=1}^{n} x_i} \exp - \frac{1}{2\delta^2} \sum_{i=1}^{n} (\log x_i - \mu)^2 \} \qquad (2.6.1.14)$$

By the usual way of finding the maximum likelihood estimators

of μ and δ^2, it can be shown that the estimators of μ and δ^2 denoting by $\hat{\theta}_1$ and s_1^2, respectively to be

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} \log X_i \qquad (2.6.1.15)$$

$$s_1^2 = \frac{1}{n} \sum_{i=1}^{n} \log (X_i - \hat{\theta}_1)^2 \qquad (2.6.1.16)$$

$$= \frac{n-1}{n} v_y^2$$

where $$v_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$y_i = \log X_i$$

The estimator s_1^2 is biased but consistent; but if equation (2.6.1.16) is replaced by $s_1^2 = v_y^2$, then, $\hat{\theta}_1$ & s_1^2 are minimum variance unbiased estimator of u and δ^2. The variance of $\hat{\theta}_1$ and s_1^2 required in determining the large sample efficiencies of other estimators are readly obtained from normal theory as

$$D^2(\hat{\theta}_1) = \frac{\delta^2}{n}$$

$$D^2(s_1^2) = \frac{2\delta^4}{n-1}$$

$$= \frac{2\delta^4}{n} \qquad (2.6.1.18)$$

When observations are given in grouped form maximum likelihood estimation of u and δ^2 can be applied but with some difficulty. For detail procedures of maximum likeli-

hood estimation of μ and δ^2 reference can be made to Aitchison and Brown (1969).

The second method, method of moments estimators of μ and δ^2 , say $\hat{\theta}_2$ and s_2^2 , respectively are obtained by equating the first sample moments m_1 and m_2 with (2.6.1.2) by substituting $\hat{\theta}_2$ and s_2^2 in place of μ and δ^2 , with j=1 and 2. So it can be shown that

$$\hat{\theta}_2 = 2 \log m_1 - \tfrac{1}{2} \log m_2 \qquad (2.6.1.19)$$

$$s_2^2 = \log m_2 - 2 \log m_1 \qquad (2.6.1.20)$$

$\hat{\theta}_2^2$ and s_2^2 are consistent and their large sample variances obtained by variational method are:

$$D^2(\hat{\theta}_2) = \frac{1}{4n}(\eta^8+4\eta^6-2\eta^4+4\eta^2) \qquad (2.6.1.21)$$

$$D^2(s_2^2) = \frac{1}{n}(\eta^8+4\eta^6+2\eta^4) \qquad (2.6.1.22)$$

where $\eta^2 = e^{\delta^2}-1$

Therefore, large sample efficiencies of $\hat{\theta}_2$ and s_2^2 are

$$\text{eff}(m_2) = \frac{D^2(\hat{\theta}_1)}{D^2(\hat{\theta}_2)}$$

$$= \frac{4\delta^2}{\eta^8+4\eta^6-2\eta^4+4\eta^2} \qquad (2.6.1.23)$$

$$\text{eff}(s_2^2) = \frac{D^2(s_1^2)}{D^2(s_2^2)}$$

$$= \frac{4\delta^2}{\eta^8+4\eta^6+2\eta^4} \qquad (2.6.1.24)$$

The efficiency of the estimators declines as δ^2 increases. The method can be applied to grouped data with Sheppard's correction if there is a high contact of the distribution with the x-axis.

The quartile estimators of μ and δ^2 are obtained by equating the sample quantiles of order q_1 and q_2 $(q_1 < q_2)$ with the expression given in (2.6.1.8) after replacing μ and δ_2^2 by their estimators $\hat{\theta}_3$ and s_3^2, respectively. So, $\hat{\theta}_3$ and s_3^2 are given by the following expressions:

$$\hat{\theta}_3 = \frac{V_{q2}\log X_{q1} - V_{q1}\log X_{q2}}{V_{q2} - V_{q1}} \qquad (2.6.1.25)$$

$$s_3^2 = \frac{\log X_{q2} - \log X_{q1}}{V_{q2} - V_{q1}} \qquad (2.6.1.26)$$

The maximum efficiency attainable by the method is when quantiles are symmetrically placed i.e. $V_{1-q} = V_q = V$ Therefore, the large sample variances of $\hat{\theta}_3$ and s_3^2 with quantilies symmetrically placed are given by:

$$D^2(\hat{\theta}_3) = \frac{\Pi\delta^2 q e^{v^2}}{n} \qquad (2.6.1.27)$$

$$D^2(s_3^2) = \frac{4\Pi\delta^4(1-2q)e^{v2}}{nv^2} \qquad (2.6.1.28)$$

and hence large sample efficiencies by

$$\text{eff}(\hat{\theta}_3) = \frac{1}{\Pi q e^{v2}} \qquad (2.6.1.29)$$

$$\text{eff}(s_3^2) = \frac{v^2}{2\,q(1-2q)e^{v2}} \qquad (2.6.1.30)$$

The method of quantiles can be applied for grouped data. But the method becomes inefficient if the data are so grouped that it is necessary to choose quantiles or pairs of quantiles distant from the more efficient quantiles that are assymetrically placed. In such a case interpolation may be preferred to obtain the most efficient quartile pairs.

The graphical method of estimating μ and δ^2 is facilitated by the use of logarthmic probability paper. The theory underlying its use emerges from the relation given by experession (2.6.1.8). Taking logarithms of both sides of (2.6.1.8) we get

$$\log x_q = \delta V_q + \mu \qquad (2.6.1.31)$$

If we let $L(x)$ to denote the proportion of sample values less than or equal to x, then

$$q_i = L(X_i) \qquad (2.6.1.32)$$

and

$$y_i = \log(X_i) . \qquad (2.6.1.33)$$

then, we should expect the points (V_{qi}, y_i) to lie approximately on straight line $y = \delta V + \mu$

The same arrayof points is obtained if we plot the points $\{L(X_i), X_i\}$ with $L(x)$ on a normal probability scale and x on logarthmic scale; the purpose of the

logarithmic probability paper is thus to facilitate the plotting of the points (V_{qi} , y_i) by providing these appropriate scales so that only $(L(x_i), x_i)$ need be computed.

The application of this method require the form of the data to be given in grouped cummulative frequency table. To obtain estimators of μ and δ^2 say $\hat{\theta}_4$ and s_4^2 respectively, the following approach is said to be appropriate.

From expression (2.6.1.8) the population quantiles of order 16, 50 and 84 are given by

$$F_{16\%} = e^{\mu-\delta^2}$$

$$F_{50\%} = e^{\mu}$$

$$F_{84\%} = e^{\mu+\delta}$$

so that

$$\mu = \log F_{50\%} \qquad (2.6.1.34)$$

$$\delta = \log \{\tfrac{1}{2}(\frac{F_{50\%}}{F_{16\%}} + \frac{F_{84\%}}{F_{50\%}}) \qquad (2.6.1.35)$$

If we read from the straight line graph $y = \delta V+\mu$ the values of x corresponding to the 16, 50 and 84% points and substitute in (2.6.1.34) and (2.6.1.35) we obtain estimates $\hat{\theta}_4$ and s_4^2 of μ and δ^2, respectively.

This method is easily applicable and simoultaneously provides a test of lognormality. The efficiencies of the

estimatorsare not calculable and it is said to be less reliable than the numerical methods.

Lognormality test for the two parameter case can be carried out by the use of skewness and kurtosis test of normality for the transformed sample values. Geary (cited in Aitchison and Brown, 1963) treats a series of tests for skewness and kurtosis on the statistics $g_1(p)$ and $g_2(p)$ defined by

$$g_1(p) = \frac{s'(p) - s''(p)}{\{s(2)\}^{\frac{1}{2}p}} \qquad (3.6.1.36)$$

and

$$g_2(p) = \frac{s(p)}{\{s(2)\}^{\frac{1}{2}p}} \qquad (3.6.1.37)$$

where

$$s(p) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}|^p$$

$$s'(p) = \frac{1}{n} \sum_{y_i > \bar{y}}^{A} |y_i - \bar{y}|^p$$

$$s''(p) = \frac{1}{n} \sum_{y_i}^{\bar{y}} |y_i - \bar{y}|^p$$

where $y_i = \log x_i$

and $p \geq 0$. He concludes that, for large samples and a wide field of alternative hypothesis regarding the nature of the population $g_1(3)$ and $g_2(4)$ are the most effective test statistics; also that for sample of moderate size $g_2(>1)$ is probabily as efficient as $g_2(4)$.

X^2 goodness of fit test is a method that may be applied for all lognormal distribution. However, this test is likely to be less sensitive than Geary since it

ignores the sign and pattern of the differences between observed and expected group frequencies and often require additional grouping at the extremes of the range, Aitchison and Brown, 1963).

Another test of lognormality is the use of the inequality measures given in their bounds form. The test using bounds of inequality measures, however, even if they are not definitive test they provide some idea about the suitability of the lognormal distribution for income data.

Given lower and upper bounds of inequality measures that could be obtained from the raw data and given inequality measures derived in relation to lognormality assumption, if the latter falls betwenn the bounds of the respective inequality measures, then it is reasonable to accept the lognormal distribution as a close approximation. This idea works too for the pareto distribution which is to be discussed in the next Section.

2.7.2 Pareto distribution

"In the course of examination of the upper tail of income distribution in a number of countries pareto found a remarkably close fit to that particular two parameter distribution which now bears his name," Cowell, (1977).

The two parameter pareto distribution is given by

$$F(X;x_o,a) = 1-(\frac{x_o}{x})^a, \quad x_o>0, \quad a>0; \quad x \geq x_o \qquad (2.7.2.1)$$

and its density function by

$$f_X(x) = \frac{ax_o^a}{x^{a+1}}, \quad a>0, \quad x \geq x_o > 0 \qquad (2.7.2.2)$$

Where X_o is some minimum income level and 'a' is known as the pareto constant and as shape parameter.

The distribution posseses moment of any order provided 'a' is not less than the order of the moment. Thus, the j^{th} order moment about the origin denoting by p_j is

$$p_j = \frac{ax_o^j}{a-j}, \quad \text{for } a > j \qquad (2.7.2.3)$$

and the mean τ and variance, v^2, is

$$\tau = \frac{ax_o}{a-1}, \quad \text{for } a > 1 \qquad (2.7.2.4)$$

$$v^2 = \frac{ax_o^2(a-1)^2}{a-2}, \quad \text{for } a > 2 \qquad (2.7.2.5)$$

The pareto distribution in the examination of the income distribution fundametal importance,(Cowell,1977) remains for the following reasons. First, the functional form works well for a number of sets of data. Second, the distribution is related to a simple random process theory of income development similar to lognormal

distribution, the main difference between the two being that a device is introduced to prevent an indefinite increase in dispersion which as the effect of erecting a lower barrier income x_o which none can fall. Third, the paretian form can be shown to result from simple hypothesis about the formation of individual remuneration within bureactratic organizations. Fourth, the functional form of the pareto distribution has some remarkably convenient properties which make it useful for description of distributional problems and for some technical manuplations. Among the convenient properties the special attraction of the pareto distribution lies on a) Exact linearity of the pareto diagram which allow to obtain the proportion of the population with income less than or equal to some specified income level x. b) There is a law attached to vander Wijk's name that particularly works for the pareto distribution. i.e. take any income level x, then the average income of everyone who get x' or more is simple $\frac{a}{a-1} x'$, where a is pareto's constant. c) The pareto distribution with parameter 'a' and x_o also provide non-intersecting Lorenz Curves which are uniquelly labelled by the pareto constant 'a'. d) easy interpretation of the parameters, and e) preservation under log-linear transformation.

For the estimation of the parameters 'a' and x_0 of the pareto distribution, method of least squares, method of moments, methods of quantiles and method of maximum likelihood are available.

Rearranging the expression given by (2.7.2.1) and taking logarithm of both sides we obtain

$$\text{Log}\,[1-F(x)] = a \log x_o - a \log x \qquad (2.7.2.6)$$

Let $P = 1-F(x)$

$$Z = a \log x_o,\ a>0\ ,\ x \geq x_o$$

hence equation (2.7.2.6) can be written as

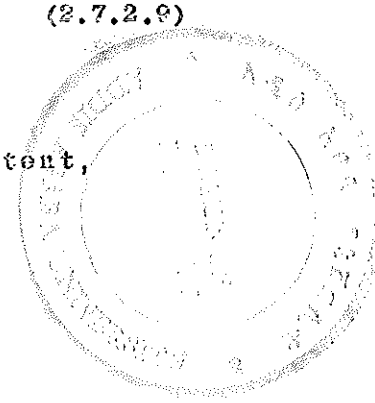
$$\text{Log}\ P = Z - a \log x$$

Expression given by (2.7.2.7) indicate that under the pareto distribution a linear relationship exists between log P and log x, so, the parameter a and x_o may be estimated by ordinary least square method from sample estimates of $F_x(x)$. Thus ordinary least square estimators of 'a' and x_o respectively are

$$\hat{a}_1 = \frac{-n\sum_{i=1}^{n} \log X_i \log P_i + (\sum_{i=1}^{n} \log X_i)(\sum_{i=1}^{n} \log P_i}{n\sum_{i=1}^{n} (\log X_i)^2 - (\sum_{i=1}^{n} \log X_i)^2} \qquad (2.7.2.8)$$

$$\hat{X}_o = \exp \frac{\frac{1}{n}(\sum_{i=1}^{n} \log P_i + \hat{a}\sum_{i=1}^{n} \log X_i)}{\hat{a}_1} \qquad (2.7.2.9)$$

Estimators obtained by this method are consistent, (Johnson & Kotz, 1970).



Method of moments also provide consistent estimators provided a > 1, the mean of the pareto distribution exists and may be obtained from the expression (2.7.2.3). Moreover, the expected value of the lowest observation, say x_1, given the assumption that the observations are from the pareto distribution can be shown to be

$$E(Y_1) = \frac{na x_o}{na-1} \quad a>0, \; x_o>0 \qquad (2.7.2.10)$$

To obtain sample estimates of a and x_o, we equate expression (2.7.2.3) with sample mean $\bar{x}$ and expression (2.7.2.10) with the lowest value x_1. Hence estimates a^* and x^* of a and x_o, respectively can be obtained from the two equation and we get

$$a^* = \frac{n\bar{X}-X_1}{n(\bar{X}-X_1)} \qquad (2.7.2.11)$$

$$x_o^* = (1-\frac{1}{a^*})\bar{X} \qquad (2.7.2.12)$$

Quantiles estimators of a and x_o are obtained by selecting two numbers q_1 and q_2 between 0 and 1 and obtain estimators of the values x_{q_1} and x_{q_1}. Then, estimators a and x_o of a and x_o, respectively may be obtained by solving two simoultaneous equations given by

$$q_1 = 1-(\frac{x_o}{x_1})^a \qquad (2.7.2.13)$$

$$q_2 = 1-(\frac{x_o}{\hat{x}_{q_2}})^a \qquad (2.7.2.14)$$

The estimator of a is then

$$\tilde{a} = \frac{\log\left(\frac{1-q_1}{1-q_2}\right)}{\log\left(\hat{X}_{q_1}/\hat{X}_{q_2}\right)} \qquad (2.7.2.15)$$

The corresponding estimator of x_o can be obtained from either of the expressions (2.7.2.13) or (2.7.2.14).

The likelihood function for a sample $(x_1, x_2, \ldots x_n)$ from a pareto distribution is

$$L = \prod_{i=1}^{n} \frac{a x_o^a}{x_j^{a+1}} \qquad (2.7.2.16)$$

Taking logarithms of both sides and differentiating partially with respect to the parameter 'a' and setting the result to zero we find the relation

$$\bar{a} = n \sum_{i=1}^{n} [\log x_i/\hat{X}_o]^{-1} \qquad (2.7.2.17)$$

which is an estimator of 'a'.

The second equation which may be obtained by taking the first derivative of log with respect to x_o cannot yield the estimate $\bar{x}_o$ in the usual way since Log L is unbouned with regard to $\bar{x}_o$. By inspection the value of $\tilde{x}_o$ which maximize (2.7.1.16) is

$$\bar{X}_o = \min_i X_i \qquad (2.7.2.18)$$

The suitability of the pareto distribution as a close approximation to income data may be tested by the use of coefficient ofdetermination(R^2)if we are using regression. However, the use of R^2 as a criteria of satisfactory fit may be misleading when fitting a Curve to a highly skewed distribution since a close fit in the tail may mask substantial, departure elsewhere. An easy alternative method is the use of the lower and upper bounds of inequality measures as indicated when we are dealing with test of the lognormal distribution.

3. Source and an Overview of the Data

The income data that is utilized in this paper is taken from the sub-sample advance report, which is a part of rural households income, Consumption and expenditure survey conducted by Central Statistics Authority collected from May 1981 to April 1982. The data are collected on agricultural sedentary population of Ethiopia excluding Tigrai and Eritrea on a household basis and the method of data collection employed was personal interviews and objective measurements.

The main sample survey on the rural income, Consumption and expenditure covers 12,000 households. However, there is a delay in reporting the main sample survey results. The main cause for the delay of the main report, as indicated by the authority, is lack of time since 100% manual editing, coding and verification of the questionnaire was made at head office. Thus, priority was given to the sub-sample advance report in editing, coding, verification and data entry. This sub-sample report consitutes 25% or 3082 households and informations are limited only to national level. The main sample survey report is expected to provide information on regional and national level. Data of the sub-sample are given in Appendix A.

The data on income is collected according to the following concepts and definitions. (Taken from the advance report, of bulletin 61, CSA, 1988).

Household: Consitutes of a person or group of persons irrespective of whether related or not, who normally live together in the same housing unit or group of housing units and who have common cooking arrangements.

Member of Household: Person consituting a house is member of the household.

The following are considered as members of a household.

a) All persons who lived with the household for at least six months including those who were not with the household at the time of the survey and were expected to be absent for less than six months.

b) All guests and visitors who stayed with household for six months and above.

c) Servants, guards, baby-sitters etc who lived with the household even for less than six months.

Income: refers to domestic consumption of own crops and own livestock and livestock products, domestic consumption of goods and services purchased for resale or produced or processed in the household enterprise other than agriculture, wages and salaries allowance, overtime, bonus, pension, commission, discounts, (i.e. concessions

obtained). Imputed rent on free housing (provided by employer), imputed rent of employer subsidized housing (i.e. subsidized amount only), other employers benefit, interest received and devidend received, imputed rent of owner occupied housing, remittance (regularly received), value of items obtained free, rent of personal possessions, alimony (regularly received) and other type of income.

The average household size for the rural sector obtained from the sub-sample is 5 persons and average income of a household is Birr 1,680.93. The highest percentage of the household, about 11.1 percent, fall in income group Birr 900-1099. Regarding the distribution of households along with their income 48 percent of the households have income less than Birr 1300 while the next 45 percent fall in income group of Birr 1300-3499. The rest 7 percent earn income above Birr 3499. Accordingly when dealing with their income share, households earning income below Birr 1300, i.e. 48 percent of the households, share 23.5 percent of total income while the next 45 percent, those who fall in income group 1300-3499 share 55.5 percent of total income. The rest 7 percent share 21 percent of total income.

Cummulative percentage distribution of households by income group and their respective cummulative income share is given in appendix B.

4. Results

The quality of the data depends on many activities which ranges from the preparation stage to the final report. The data that has been utilized in this paper which was published by CSA, (1988) has undergone all the necessary stages. The data was collected by well trained enumerators and under careful supervision to meet many objectives at a national level. Moreover, important parts of data processing, such as editing, coding, verification data entry and tabulation was done at the head offfice of the Authority so as to ensure the quality of the data.

From the sub-sample data average household size of the rural population was 5 persons which is slightly higher over the 1981 demographic survey which gave 4.8 persons per household. Moreover, per capita income from the sub-sample data was Birr 333.- which seems also high. The overestimation could be due to the exclusion of Tigrai and Eritrea in the sub-sample. These two regions are believed to have mean income which is below the national average income.

4.1. Inequality Measures

The inequality measures considered in this paper are given in their lower and upper bound forms. An attempt was made to estimate the compromise values of the bounds by fitting the pareto distribution in each income group. But the inavailability of appropriate computer soft ware has been a bottleneck. However, for the most commonly used inequality measure, Gini coefficient, the compromise value can be approximated by the expression given in (2.2.4).

In most cases income data given in grouped form are open ended. The sub-sample income data utilized in this paper also reveal this nature. From section 2.5 it can be seen

that the estimation of the upper bound of the inequality measures require all class limits to be known. Therefore, it is assumed that the lowest income of households of the sub-sample to be Birr 1.00 and the highest income to be Birr 14,500. Thus, the inequality measures lower and upper bound estimates are confined to this income range.

Table 4.1.1. shows the lower and upper bounds of four inequality measures and the difference between the bounds.

Table 4.1.1. Lower and Upper bounds of inequality measures

Inequality measures	Lower bound	Upper bound	Difference
Gini Coefficient	0.3624	0.3654	0.003
Theil's index	0.2292	0.2319	0.0027
Relative mean deviation	0.5215	0.5225	0.001
Logarthmic variance	0.5187	0.6487	0.1300

The compromise value for the Gini coefficient is found to be 0.3644.

Gini coefficient and Theil's entropy index show more equality compared to Relative mean deviation and Logarthmic variance. This is true for lower and upper bounds. Also, except for the logarthmic variance the other three measures are quite stable. This is shown by comparing the differences between the upper and lower bounds of each measures.

4.2. Results of the Lognormal Distribution

For the estimation of u and σ^2 of the lognormal distribution two methods are employed. These are methods of moment and method of quantiles at quantiles of order 27% and 73%. Accordingly three methods of testing the lognormality assumption as a close fit of the sub-sample income data have been adopted. These include, skewness and kurtosis tests of normality for the transformed data, the x^2 - goodness of fit test and tests using the lower and upper bounds of the inequality measures. But all tests resulted in the lognormal distribution to be a poor fit to the sub-sample income data.

both cases x^2 values calculated from the fit are greater than table value indicating the lognormal distribution as a close fit to the income data under consideration is invalid. As it can be seen from Appendix C the major contributors for the departure are in the low income brackets and to some extent in the upper income brackets.

Tests using the lower and upper bounds of inequality measures has also resulted in the same conclusion in rejecting the lognormality assumption. Table 4.2.2 shows inequality measures derived in relation to lognormality assumption and the lower and upper bounds without any assumption of distribution.

Table 4.2.2. Estimates of inequality measures with log-normality assumption and without assumption

Inequality measures	With Lognormality Assumption	Without Assumption	
		Lower bound	Upper bound
Gini coefficient			
method of moments	0.3688	0.3624	0.3654
method of quantiles	0.3616		
Theil's index			
method of moments	0.2349	0.2292	0.2319
method of quantiles	0.2225		
Relative mean deviation			
method of moments	0.5324	0.5215	0.5225
method of quantiles	0.5172		
Logarthmic variance			
method of moments	0.5249	0.5187	0.6478
method of quantiles	0.4945		

With the exception of the logarthmic variance, the estimates obtained by lognormality assumption lie beyond the upper bound for method of moments while with method of quantiles they lie below the lower bound. This has happened because of the difference between the variance δ^2, estimates obtained by the two methods. In other words, inequality measure estimates derived with the assumption of lognormality depend on the value of δ^2

4.3. Results of the Pareto distribution

The Pareto constant 'a' is estimated using ordinary least square method which works both for ungrouped and grouped data case. For the whole distribution pareto distribution is found to be a poor fit with pareto's 'a' equal to 0.54 and coefficient of determination, R^2, equal to 0.4 which is low. The test using lower and upper bounds is not also infavour of the pareto distribution to be a good fit for the income data as a whole. Infact, the pareto distribution is said to be a close approximation to the upper tail of income data. Hence a consecutive fits and tests using the lower and upper bounds of inequality measures along with R^2, pareto distribution is found to be a good fit for groups having income of Birr 1700.- and above i.e. for the last ten income groups. For these 10 groups the fit has resulted in 'a' to be 2.7563 and R^2 to be 0.99.

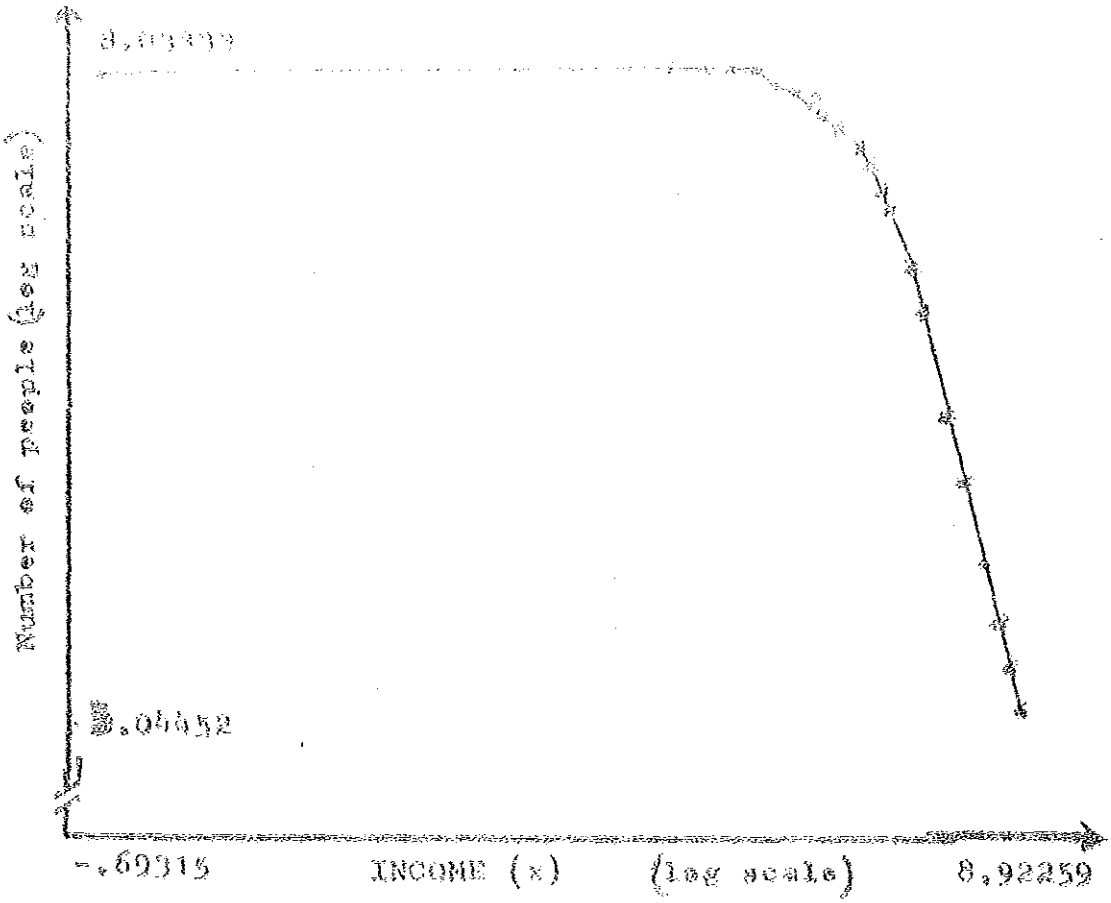


Fig. 4.3.1 Number of people receiving income above or equal to an income level x Vs income level x on logarithmic scale.

For convenedency Fig. 4.3.1 is plotted by using number of person, in logarithmic scale, rather than proportions. In fact, both yield same result to depict the relationship under consideration.

As it is seen from Fig. 4.3.1 the exact linearity of the pareto distribution seem to work for the last ten income groups.

Table 4.3.1. shows inequality measures estimates with paretian assumption and bounds without assumption for the last ten income groups.

Table 4.3.1. Inequality measures for the last 10 income group

Inequality measures	with assumption	Without assumption	
		Lower bound	Upper bound
Gini Coefficient	0.2216	0.2180	0.2261
Theil index	0.1136	0.0958	0.1190
Rel. mean dev.	0.3288	0.3215	0.3292
Logarthmic variance	0.1393	0.1404	0.1501

Except for logarthmic variance, inequality measures derived in relation to pareto distribution, lie between the lower and upper bounds. Since three of the inequality measures satisfy the criteria with R^2 equal 0.99 it may be argued that the pareto distribution to work well for the last 10 income groups.

5. Discussion and Conclusion

The results of this paper are based on the sub-sample income data which is a part of the main rural households income, consumption and expenditure survey conducted by Central Statistics Authority (CSA) covering a period of one year from May 1981-April 1982. The data are collected on the agricultural sedentary population of Ethiopia excluding Eritrea and Tigrai on a household basis.

As mentioned in Chapter 3, the sub-sample data constitutes about 3082 households which is about 25% of the total sample survey whose coverage is about 12,000 households. Therefore, since results of the sub-sample might differ from the main survey, we might be able to take the results in this paper as informative rather than conclusive.

The lower and upper bounds of inequality measures show differences which are small. These small differences are attributed to grouping effects (see Table 4.1.1).

The lognormal distribution has failed to represent the whole distribution of the sub-sample income data. As it can be seen from Table 4.2.1. relative measures of skewness and kurtois that are utilized for the tes$t of normality for the transformed data, when compared

with respective table values show a slight departure of the transformed data from normality, and hence a departure from lognormality of the untransformed data. Tests using the lower and upper bounds (Table 4.2.2.) are also in support of the above test results with minor departure of the lognormal distribution as a close approximation of the whole income data. Moreover, according to the x^2 goodness of fit test as it may be seen from the table in appendix C, the major contributors of the departure are the lower income brackets and to some extent the upper income brackets. This result is contrary to the existing theory of the lognormal distribution as a better approximation to the low income range.

The pareto distribution has been found to be a good fit for groups having income Birr 1700.- and above. This result is supported by a test using the lower and upper bounds (see Table 4.3.1) along with R^2 equal 0.99. Moreover, from Fig. 4.3.1 it can be seen that the pareto distribution exactlinearity works will for the last 10 income groups or groups having income Birr 1700.- and above. Therefore it can be argued that the pareto distribution is incomformity with the existing theory of its close approximation to the upper tail of income data.

Therefore, from the above discussion it might be possible to consider some conclusive remarks to the sub-sample rural Ethiopia households income.

The difference obtained between the upper and lower bounds of inequality measures are small. These small differences indicate minor effects attributed to grouping.

The contrary result of the lognormal distribution as a close fit to the low income range might be attributed to sub-sampling error and grouping effects. According to x^2-goodness of fit test, it may be argued that the lognormal distribution to be a better approximation to the middle income brackets.

The Exact linearity of the pareto diagram works well for the last ten income groups. This result is in agruement to the pareto distribution to be a close fit to the upper tail of income data.

References

Aitchison, J. and Brown, J.A.C. (1963). "The Lognormal distribution with special reference to its uses in economics." Cambridge University Press, UK.

Cowell F.A. (1977). "Measuring Inequality." Phillip Allan Publishers LTD, Deddungton, Oxford

Croxton Frederick E. and Cowden Dudly J. (1960). "Applied general statistics." Printice - Hall inc. USA

CSA (1988). "Rural Household Income, Consumption and expenditure survey." Bulletin No. 61 Addis Ababa

Johnson Norman L. and Kotz Samuel (1970). "Continuous Univeriate distribution - 1. " John Wiley & Sons,USA

Kuznets, S. (1955). " Economic growth and Income Inequality." American Economic Review, 45, pp 1-23

Kuznets, S. (1963). Quantitative Aspects of Economic Growth of Nations: Distribution of income by size. "Economic development and cultural Change, vol. xi, No. 2, part II, January, pp 1-8

Mann Arthur J. (1984). " Economic Development, Income distribution and income levels." Puerto Rice 1953-1975." Economic development and cultural change, vol. 33, No.3, April, pp 485-502

Paglin M. (1975) " The measurement and trend of inequality, A basic revision" American economic review, Sept.

Sundrum, R.M. (1974). " Aspects of Economic Inequality in Developing Countries." The Bangladesh Development Studies, vol. II, No.1, January, pp 445-466

Appendix A: percentage distribution of households by income group.

Income Group (Birr)	Percentage of households	Group mean (Birr)
Below 199	0.81	144.41
200 - 299	1.20	244.84
300 - 399	2.17	357.10
400 - 499	2.73	452.11
500 - 599	3.76	550.02
600 - 699	5.32	648.82
700 - 799	5.48	750.98
800 - 899	5.39	847.77
900 - 1099	11.10	999.38
1100 - 1299	9.86	1193.13
1300 - 1499	7.98	1391.13
1500 - 1699	7.59	1598.97
1700 - 1899	6.29	1797.46
1900 - 2299	9.77	2086.95
2300 - 2699	6.13	2488.30
2700 - 3499	7.62	3046.82
3500 - 4299	2.92	3850.96
4300 - 5099	1.75	4710.78
5100 - 5899	0.75	5377.57
5900 - 6699	0.42	6199.19
6700 - 7499	0.26	7047.73
7500 & above	0.68	10593.91

Appendix B: Cummulative percentage distribntion of households and cummulative percentage income share by income group.

Income group (Birr)	Cummulative percentage of households	Cummulative percentage income share
below 199	0.81	0.07
200 - 299	2.01	0.25
300 - 399	4.18	0.71
400 - 499	6.91	1.44
500 - 599	10.67	2.67
600 - 699	15.99	4.72
700 - 799	21.47	7.17
800 - 899	26.86	9.89
900 -1099	37.96	16.48
1100 -1299	47.82	23.48
1300 -1499	55.80	30.08
1500 -1699	63.39	37.30
1700 -1899	69.68	44.02
1900 -2099	79.45	56.15
2300 -2699	85.58	65.22
2700 -3499	93.20	79.02
3500 -4299	96.12	85.71
4300 -5099	97.87	90.61
5100 -5899	98.62	93.08
5900 -6699	99.04	94.62
6700 -7499	99.30	95.71
7500 & above	100.00	100.00

Appendix C: X^2 values calculated from the lognormal distribution fit.

Income Group	Observed frequencies N=3082	Method of Moments		Method of Quantiles	
		Expected frequencies	$\frac{(O-E)^2}{E}$	Expected frequencies	$\frac{(O-E)^2}{E}$
Below 199	25.0	8.6	31.27	6.5	52.65
200 - 299	37.0	37.6	0.01	30.2	1.53
300 - 399	66.9	77.4	1.42	66.9	-
400 - 499	84.1	111.9	6.91	102.3	3.24
500 - 599	115.9	143.6	5.34	136.8	3.19
600 - 699	164.0	155.9	0.42	153.8	0.68
700 - 799	168.9	172.6	0.08	165.5	0.07
800 - 899	166.1	168.6	0.04	172.9	0.27
900 - 1099	342.1	324.8	0.92	331.0	0.37
1100 - 1299	303.9	303.0	0.00	301.4	0.02
1300 - 1499	245.9	257.0	0.48	257.4	0.51
1500 - 1699	233.9	213.3	1.99	226.5	0.24
1700 - 1899	193.9	178.4	1.35	190.8	0.05
1900 - 2299	301.1	276.1	2.26	278.0	1.92
2300 - 2699	188.9	186.2	0.04	195.1	0.20
2700 - 3499	234.8	222.5	0.68	227.1	0.26
3500 - 4299	90.0	110.0	3.64	111.0	3.97
4300 - 5099	53.9	57.3	0.20	57.0	0.17
5100 - 5899	23.8	30.8	1.59	29.9	1.24
5900 - 6699	12.9	18.2	1.54	16.6	0.82
6700 - 7499	8.0	10.0	0.40	9.3	0.18
7500 & above	21.0	18.2	0.43	16.0	1.56
Total	3082	3082	61.02	3082	73.14

Appendix D: X^2 - values calculated for pareto distribution fit.

Income Group (Birr)	Observed frequencies (O)	Expected frequencies (E)	$\frac{(O-E)^2}{E}$
less than 199	25	1396.1	14,203.9
200 - 299	37	335.9	265.0
300 - 399	66.9	194.2	83.4
400 - 499	84.1	132.5	17.7
500 - 599	115.9	95.5	4.4
600 - 699	164.0	74.0	109.5
700 - 799	168.9	58.6	207.6
800 - 899	166.1	49.3	276.7
900. - 1099	342.1	77.1	910.8
1100 - 1299	303.9	58.6	1026.8
1300 - 1499	245.9	43.1	954.2
1500 - 1699	233.9	37.0	1047.8
1700 - 1899	193.9	30.8	863.7
1900 - 2299	301.1	49.3	1286.1
2300 - 2699	188.9	37.0	623.6
2700 - 3499	234.8	55.5	579.3
3500 - 4299	90.0	37.0	75.9
4300 - 5099	53.9	27.7	24.8
5100 - 5899	23.8	21.6	0.22
5900 - 6699	12.9	18.5	1.70
6700 - 7499	8.0	12.3	1.50
7500 & above	21.0	240.4	200.2
Total	3082	3082	22,764.82