A New Functional Form
For Estimating Lorenz Curves

Sharif Hossain, (E-mail: sharif@en.kyushu-u.ac.jp), Kyushu University, Japan

Chikayoshi Saeki, (E-mail: saeki@en.kyushu-u.ac.jp), Kyushu University, Japan

Abstract

The Lorenz curve is a most powerful tool in the analysis of the size distribution of income and wealth. In the past decades, many authors have proposed different functional forms for estimating Lorenz curves from grouped data. Most of the functional forms do not fit the data very well for estimating Lorenz curves. That is why, in this paper we proposed a new functional form for estimating Lorenz curves, which provides very good fits with compared to other functional forms, see for example, Kakwani's and Podder's (1973, 76), Rasche's et all (1980), Kakwani's (1980), Gupta's (1984), and Ortega's (1991) functional forms. On the basis of the new functional form, we derived the formulae of the Gini, Kakwani, and Chakravarty inequality indices. Empirical verification of the theoretical construct has been done based on the data set from BBS (Bangladesh Bureau of Statistics) publications, "Household Expenditure Survey" corresponding to different years.

1. Introduction

*T*he Lorenz curve is defined as the relationship between the cumulative proportion of income units and the cumulative proportion of income received by these units. Let, p(x) is the proportion of units receive income up to x, and q(x) is the proportion of total income received by the same units. Then the Lorenz curve is the graphical representation of the parametric relationship between p and q. The graph of the curve is represented in a unit square. The straight line joining the points (0, 0) and (1, 1) is called the egalitarian line, because along this line p = q, which means that each unit receive the same income. The Lorenz curve falls below the egalitarian line. When two or more than two Lorenz curves coincide two each other, we cannot say which one is better than other. Thus, to overcome this type of problem, in the past decades, many authors have proposed different functional forms for estimating Lorenz curves from grouped data, see for example, Kakwani and Podder (1973, 76), Rasche et all (1980), Kakwani (1980), Gupta (1984), and Ortega (1991). Most of these functional forms do not fit the data very well. That is why, in this paper we proposed a new functional form for estimating Lorenz curves based on the studies of the income distribution of Bangladesh, which provides very good fits with compared to Kakwani's and Podder's (1973, 76), Rasche's et all (1980), Kakwani's (1980), Gupta's (1984), and Ortega's (1991) functional forms. For measuring income inequality, several measures have been developed, see for example Gini's (1913) measure, Theil (1967), Atkinson (1970), Kakwani (1980), Basman and Slottje (1987, 1988) and Chakravarty (1988) have all formulated different measures.

———————————

Readers with comments or questions are encouraged to contact the authors via email.

Among them the most widely used measure is the Gini's concentration ratio. It can be shown that the concentration ratio (suggested by Gini) is equal to twice the area between the Lorenz curve and the egalitarian line. The lower and upper limits of the concentration ratio are zero and unity respectively. If each unit receives the same income, obviously the Lorenz curve coincides with the egalitarian line and the concentration ratio is zero. If on the other hand, a single unit receives all incomes the concentration ratio becomes unity. In this paper, on the basis of the new functional form we provided the formulae of the Gini, Kakwani and Chakravarty inequality measures and then we computed them. The outline of this paper is as follows;

In section 2, we discussed about the characteristics of the Lorenz curve. In section 3, we presented the new functional form for estimating Lorenz curves. In section 4, we have shown that the new functional form satisfies all the required properties of the Lorenz curve. In section 5, in order to test whether the parameters that are included in the new functional are statistically significant or not, we specified various Lorenz curve hypotheses. In section 6, on the basis of the new functional form we derived the formulae of some inequality measures say; Gini, Kakwani, and Chakravarty inequality measures. In section 7, we discussed about the data collection for empirical analysis. In section 8, an empirical analysis has been done on the basis of the data set from BBS (Bangladesh Bureau of Statistics) publications "Household Expenditure Survey" corresponding to different years. Finally, in section 9, an overall discussion and conclusion has been provided.

Next, we will discuss about the characteristics of the Lorenz Curve.

2. A Characterization of the Lorenz Curve

Lorenz (1905) was the first to present the graphical relationship between the cumulative distribution of income units (earners) ordered by income and the size distribution of income by the same units. Mathematically, the Lorenz curve is defined as follows;

Suppose income X of a unit is a random variable with the probability density function f(x). Then the function F(x) is defined as:

$$F(x) = \text{Pr } ob(X \leq x) = \int_0^x f(x)dx \tag{1}$$

where, F(x) can be interpreted as the proportion of units having an income less than or equal to x. F(x) obviously varies from 0 to 1. Further if it is assumed that the mean income μ of the distribution exists, which is given by:

$$\mu = \int f(x)dx \tag{2}$$

Then the first moment distribution function of X is defined as:

$$F_1(x) = \frac{1}{\mu} \int_0^x Xf(x)dx \tag{3}$$

where, $F_1(x)$ also varies from 0 to 1. It follows that $F_1(x)$ is interpreted as the proportional share of the total income of

the units having an income less than or equal to x. Then the Lorenz curve is the relationship between the variables F(x)

and $F_1(x)$ and it can be obtained by inverting functions (1) and (3) and eliminating x, if the functions are conveniently

invertible. Alternatively, the curve can be plotted by generating the values of F(x) and $L^{I}(x)$ from (1) and (3) by considering the arbitrary values of x. The curve is represented in a unit square. Gastwirth (1971) define the Lorenz curve

$$L(z) = \frac{1}{\mu} \int_0^z F^{-1}(x)dx; \tag{4}$$

L(z) is interpreted as the share of income of the ith class and z is the proportion of the income units of the ith class. As for example, if we looked at (say) quintiles and the 30 percent of the population had 10 percent of the income, then for the year in question L(z) would be (0.10) and z would be (0.30). This method obviously exhibits increasing degrees of freedom as the income class quintiles increase.

Kakwani (1980) noted that, the Lorenz curve L(z) should exhibits the following properties:

(i)   L(z) = 0; if z = 0   (5)
(ii)   L(z) = 1; if z = 1   (6)

(iii) $L'(z) \geq 0$; for $0 \leq z \leq 1$ (7)

(iv) $L''(z) \geq 0$; for $0 \leq z \leq 1$ (8)

(v) $L(z) \leq z$; for $0 \leq z \leq 1$ (9)

In addition Kakwani (1980a) has noted a number of other properties that the Lorenz curve possesses. He lists these as lemmas in his comprehensive discussion of the Lorenz curve in his 1980 book. We present these lemmas without comment as follows:

1.   The distance between the Lorenz curve and the egalitarian line is a maximum at income level X = μ

2.   Dividing the population into two groups so in the first group all the income units have income less than μ , the proportion of income that should be transferred so both groups have the same income is given by the maximum distance between the Lorenz curve and the egalitarian line.

3.   The Lorenz curve q = L(z) is symmetric iff 1-z = L(1-q).

4.   If the Lorenz curve q = L(z) is symmetric, the point $(z_\mu, L(z_\mu))$ corresponding to mean income μ lies on the diagonal perpendicular to the egalitarian line.

5.   The necessary and sufficient condition for the Lorenz curve to be symmetric is $\frac{f(\mu^2/x)}{f(x)} = \left(\frac{x}{z}\right)^3$ for a density f(x) for all X.

6.   The Lorenz curve for the Log-normal distribution is symmetric.

7.   The Lorenz curve q = L(z) is skewed toward (0, 0) iff $z_\mu + L(z_\mu) > 1$

8.   The Lorenz curve for the Pareto distribution is skewed toward (0, 0).

The interested reader is encouraged to review Kakwani (1980a) for a complete discussion of these properties.

Next, we will discuss about the new functional form for estimating Lorenz curves.

3. A New Functional Form for Estimating Lorenz Curves

In the past decades, many authors have proposed different functional forms for estimating Lorenz curves from grouped data. From our point of view the more relevant are as follows:

$$L(z) = ze^{\beta(z-1)}; \quad \beta > 0, \ 0 \le z \le 1 \tag{10}$$

$$L(z) = z^\alpha e^{\beta(z-1)}; \quad \alpha > 0, \ \beta > 0, \ 0 \le z \le 1 \tag{11}$$

$$L(z) = [1-(1-z)^\alpha]^{\frac{1}{\beta}}; \quad \alpha > 0, \ 0 < \beta \le 1, \ 0 \le z \le 1 \tag{12}$$

equations (10) and (11) are proposed by Kakwani and Podder (Kakwani and Podder, 1973, 1976), and equation (12) is proposed by Rasche et all (Rasche, Gaffney, Koo and Obst, 1980) respectively. Other well-known proposed functional forms for estimating Lorenz curves are as follows;

$$L(z) = zA^{(z-1)}; \quad A > 0, \ 0 \le z \le 1 \tag{13}$$

$$L(z) = z - Az^\alpha(1-z)^\beta; \quad A > 0, \ \alpha > 0, \ 0 < \beta \le 1, \ 0 \le z \le 1 \tag{14}$$

due to Gupta (Gupta, 1984) and Kakwani, (Kakwani 1980) respectively. However, in many cases the functional form (14) can not be used successfully as a Lorenz curve because it is not positive in all cases. Functional forms (11) and (12) have been successfully used in (Fernandez Morales et al, 1989; Garcia Lizana et al 1989) to estimate the Lorenz curve of the income distribution of Spanish provinces and then to compute the Gini's inequality index and several poverty indices. Another functional form is proposed by Ortega (Ortega, G Martin, A. Fernandez M Ladoux and A. Garcia, 1991), which is as follows:

$$L(z) = z^\alpha[1-(1-z)^\beta]; \quad \alpha > 0, \ 0 < \beta \le 1, \ 0 \le z \le 1 \tag{15}$$

The problem is that, most of these functional forms do not fit the data very well. That is why, in this paper, we proposed a new functional form for estimating Lorenz curves, which fits the data very well with compared to Kakwani's and Podder's (1973, 1976), Rasche's et all (1980), Kakwani's (1980), Gupta's (1984), and Ortega's (1991) functional forms.
We suggest the following functional form for estimating Lorenz curves from grouped data;

$$L(z) = z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]; \quad \alpha > 0, \ \beta > 0, 0 < \theta \le 1, \ 0 \le z \le 1 \tag{16}$$

From the properties of the Lorenz curve it has been well known that the function L(z) of the equation (16), represents the Lorenz curve, if it satisfies the properties from (i) to (v) in page (3).

Next, we will move to prove that our proposed functional form satisfies all the required properties of the Lorenz curve.

4. Proof of Properties of the Lorenz Curve for the Considered Functional Form

Property (i): From equation (16) we see that when z = 0, then L(z) is also 0, means that, L(0) = 0. So, property (i) is satisfied.

Property (ii): when z = 1, then L(z) is equal to 1, means that, L(1) = 1. So property (ii) is satisfied.

Property (iii): Now taking differentiation of the equation (16) with respect to z, then we have:

$$L'(z) = \alpha z^{(\alpha-1)} e^{\beta(z-1)}[1-(1-z)^{\theta}] + \beta z^{\alpha} e^{\beta(z-1)}[1-(1-z)^{\theta}] + \theta z^{\alpha} e^{\beta(z-1)}[(1-z)^{(\theta-1)}] \tag{17}$$

which implies that $L'(z) \geq 0$ for $0 \leq z \leq 1$. So, property (iii) is satisfied.

Property (iv) : Now again taking differentiation of the equation (17) with respect to z then we have:

$$L''(z) = \alpha(\alpha-1)z^{(\alpha-2)} e^{\beta(z-1)}[1-(1-z)^{\theta}] + 2\alpha\beta z^{(\alpha-1)} e^{\beta(z-1)}[1-(1-z)^{\theta}] + 2\alpha\theta z^{(\alpha-1)} e^{\beta(z-1)}(1-z)^{(\theta-1)} +$$

$$\beta^2 z^{\alpha} e^{\beta(z-1)}[1-(1-z)^{\theta}] + 2\alpha\beta z^{\alpha} e^{\beta(z-1)}(1-z)^{(\theta-1)} + \theta(1-\theta)e^{b(z-1)}(1-z)^{(\theta-2)} \tag{18}$$

From equation (18) it is clear that $L''(z) \geq 0$ for $0 \leq z \leq 1$, and that $L''(z) \geq 0$ if $\alpha > 1$. Now we will show

$L''(z) \geq 0$ that if $\alpha < 1$. Let us define that:

$$u(z) = -\alpha(1-\alpha)z^{(\alpha-2)} e^{\beta(z-1)}[1-(1-z)^{\theta}] \text{ and } v(z) = 2\alpha\theta z^{(\alpha-1)} e^{\beta(z-1)}(1-z)^{(\theta-1)} \tag{19}$$

Now dividing in both u(z) and v(z) in equation (19) by $z^{(\alpha-2)} e^{\beta(z-1)}$ then we have:

$$u(z) = -\alpha(1-\alpha)[1-(1-z)^{\theta}] \text{ and } v(z) = 2\alpha\theta z(1-z)^{(\theta-1)} \tag{20}$$

Thus to prove $L''(z) \geq 0$ in (0, 1) is equivalent to prove $u(z) \leq v(z)$ in (0, 1). To do this, let's compute

$$u'(z) = -\alpha\theta(1-\alpha)(1-z)^{(\theta-1)} \text{ and } v'(z) = 2\alpha\theta(1-z)^{(\theta-1)} + 2\alpha\theta(1-\theta)(1-z)^{(\theta-1)} \tag{21}$$

Since, $u'(z)$ is clearly smaller than the first term of $v'(z)$ if $\alpha < 1$, and the second term of $v'(z)$ is non-negative in

(0,1) we obtained $u'(z) \leq v'(z)$ for z in (0,1). This relation together with u(0)= v(0) = 0, assures that u(z) < v(z) for

every z in (0,1). Thus $L''(z) \geq 0$ if $\alpha < 1$. Therefore the property (iv) is satisfied.

Property (v): The equation (16) can be rewritten as follows:

$$L(z) = \frac{z^{\alpha}}{e^{\beta(z-1)}}[1-(1-z)^{\theta}] \tag{22}$$

which implies that L(z) < z for 0 < z <1. Therefore the property (v) is satisfied. Since the functional form,

$L(z) = z^{\alpha} e^{\beta(z-1)}[1-(1-z)^{\theta}]$ is satisfied all the required properties of the Lorenz curve. Therefore, our considered

functional form is of the Lorenz curve. Functional form (16) includes as a particular case when, $\alpha = 0$ and $\beta = 1$, then

the Lorenz curve will be the Gamma function. In addition when $\alpha = 1$, $\beta = 0$, and $\theta = 0$, then we obtain the egalitarian line. In order to test whether the parameters that are included in the new functional form are statistically significant or not we can specify the following null hypotheses;

5. Specification of Various Lorenz Curve Hypotheses from the New Functional Form

$$H_0^1 : \alpha = 0, \quad L(z) = e^{\beta(z-1)}[1-(1-z)^\theta] \tag{23}$$

$$H_0^2 : \beta = 0, \quad L(z) = z^\alpha[1-(1-z)^\theta] \tag{24}$$

$$H_0^3 : \theta = 1, \quad L(z) = z^{\alpha+1}e^{\beta(z-1)} \tag{25}$$

$$H_0^4 : \alpha = 0, \beta = 0, \quad L(z) = [1-(1-z)^\theta] \tag{26}$$

We have found that at every point in the sample L(z) satisfied properties (i) – (v). Imposing the restriction implied by H_0^1

through H_0^4, that L(z) satisfied all the properties. We can test these null hypotheses on the basis of the F-test statistic. All

of these specifications are of course subject to empirical examination as to their validity. Actual estimation of the descriptive approximations of the Lorenz curves given in (23)-(26) can be estimated by the non-linear least squared method.

Next, we will derive some inequality measures on the basis of the new functional form.

6. Some Inequality Measures on the Basis of the New Functional Form

The Gini index, Kakwani and Chakravartay inequality measures can be obtained on the basis of the new functional form as follows:

Proposition : If, $L(z) = z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]$, where, $\alpha > 0$, $\beta > 0$, and $0 < \theta \le 1$, then its Gini index (GI) is given by:

$$GI = 1 - 2\left[\frac{1}{\alpha+1} - B(\alpha+1, \theta+1) + \sum_{j=1}^{\infty}(-1)^j \frac{\beta^j}{j!}[B(\alpha+1, j+1) - B(\alpha+1, \theta+j+1)]\right] \tag{27}$$

where B is the beta function.

Proof : The Gini index (GI) is defined as:

$$GI = 1 - 2\int_0^1 L(z)d(z) = 1 - 2\int_0^1 z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]dz \tag{28}$$

Let, $I_g = \int_0^1 z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]dz \tag{29}$

Now, using the expansion of the power series $z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]$, then the equation (29) can be written as follows:

$$I_g = \int_0^1 z^\alpha dz + \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!} \int_0^1 z^\alpha (1-z)^j dz - \int_0^1 z^\alpha (1-z)^\theta dz - \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!} \int_0^1 z^\alpha (1-z)^{(\theta+j)} dz$$

$$\Rightarrow I_g = \frac{1}{\alpha+1} + \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!} B(\alpha+1, j+1) - B(\alpha+1, \theta+1) - \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!} B(\alpha+1, \theta+j+1)$$

$$= \frac{1}{\alpha+1} - B(\alpha+1, j+1) + \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!}[B(\alpha+1, j+1) - B(\alpha+1, \theta+j+1)] \tag{30}$$

where B is the beta function. Therefore, the Gini index is given by;

$$GI = 1 - 2\left[\frac{1}{\alpha+1} - B(\alpha+1, \theta+1) + \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!}[B(\alpha+1, j+1) - B(\alpha+1, \theta+j+1)]\right] \tag{31}$$

Another advantage of the equation (16) is the possibility of generating easy formulation for inequality measures associated with the Lorenz curve. Kakwani (1980) introduced the inequality measure K_r is defined as;

$$K_r = 1 - r(r+1)\int_0^1 L(z)(1-z)^{(r-1)} d(z) \tag{32}$$

where, L(z) is the Lorenz curve. Now from our functional form we will obtain the Kakwani inequality measure.

Proposition : If, $L(z) = z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]$, where, $\alpha > 0$, $\beta > 0$, and $0 < \theta \le 1$, then the Kakwani inequality measure K_r, is given by:

$$K_r = 1 - r(r+1)\left[B(\alpha+1, r) - B(\alpha+1, \theta+r) + \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!}[B(\alpha+1, j+r) - B(\alpha+1, \theta+j+r)]\right] \tag{33}$$

where B is the beta function.

Proof : Let us define;

$$I_k = \int_0^1 L(z)(1-z)^{(r-1)} dz = \int_0^1 z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta](1-z)^{(r-1)} dz$$

$$= \int_0^1 z^\alpha e^{\beta(z-1)}(1-z)^{(r-1)} dz - \int_0^1 z^\alpha e^{\beta(z-1)}(1-z)^{(\theta+r-1)} dz \tag{34}$$

Now, using the expansion of the power series $z^\alpha e^{\beta(z-1)}(1-z)^{(r-1)}$ and $z^\alpha e^{\beta(z-1)}(1-z)^{(\theta+r-1)}$, then the equation (34) can be written as follows:

$$I_k = \int_0^1 z^\alpha (1-z)^{(r-1)} dz + \sum_{j=1}^\infty (-1)^j \frac{\beta^j}{j!} \int_0^1 z^\alpha (1-z)^{(j+r-1)} dz - \int_0^1 z^\alpha (1-z)^{(\theta+r-1)} dz -$$

$$\sum_{j=1}^{\infty}(-1)^j\frac{\beta^j}{j!}\int_0^1 z^\alpha(1-z)^{(\theta+j+r-1)}dz \tag{35}$$

$$\Rightarrow I_k = B(\alpha+1,r)+\sum_{j=1}^{\infty}(-1)^j\frac{\beta^j}{j!}B(\alpha+1,j+r)-B(\alpha+1,\theta+r)-\sum_{j=1}^{\infty}(-1)^j\frac{\beta^j}{j!}B(\alpha+1,\theta+j+r) \tag{36}$$

Thus, $I_k = B(\alpha+1,r)-B(\alpha+1,\theta+r)+\sum_{j=1}^{\infty}(-1)^j\frac{\beta^j}{j!}[B(\alpha+1,j+r)-B(\alpha+1,\theta+j+r)]$ $\qquad(37)$

where B is the beta function. Therefore the Kakwani inequality measure is given by:

$$K_r = 1-r(r+1)\left[B(\alpha+1,r)-B(\alpha+1,\theta+r)+\sum_{j=1}^{\infty}(-1)^j\frac{\beta^j}{j!}[B(\alpha+1,j+r)-B(\alpha+1,\theta+j+r)]\right] \tag{38}$$

Also Chakrabarty (C_r, 1988) introduced another inequality measure which is defined as:

$$C_r = 2\left[\int_0^1(z-L(z))^r dz\right]^{1/r} \tag{39}$$

Now, on the basis of the new functional form, we also derive the formula of the Chakravarty inequality measure.

Proposition : If, $L(z)=z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]$, where, $\alpha>0$, $\beta>0$, and $0<\theta\le 1$, then the Chakravarty in-

equality measure (C_r,) is given by;

$$C_r = \left[\sum_{i=0}^{r}\sum_{k=0}^{i}\sum_{j=0}^{\infty}\frac{(i\beta)^j}{j!}(-1)^{(i+k+j)}\binom{r}{i}\binom{i}{k}B(r+i(\alpha-1)+1,k\theta+j+1)\right]^{1/r} \tag{40}$$

where B is the beta function.

Proof: Let us define:

$$I_c = \int_0^1(z-L(z))^r dz = \int_0^1\left[z-z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]\right]^r dz \tag{41}$$

Now using the expansion of the power series $\left[z-z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]\right]^r$, we can write that:

$$\left[z-z^\alpha e^{\beta(z-1)}[1-(1-z)^\theta]\right]^r = \sum_{i=0}^{r}(-1)^i\binom{r}{i}z^{r+i\alpha-i}e^{i\beta(z-1)}(1-(1-z)^\theta)^i \tag{42}$$

Again using the expansion of the power series $(1-(1-z)^\theta)^i$ then we have:

$$(1-(1-z)^\theta)^i = \sum_{k=0}^{i}(-1)^k\binom{i}{k}(1-z)^{k\theta} \tag{43}$$

Again using the expansion of the power series $e^{i\beta(z-1)}$, it can be written as:

$$e^{i\beta(z-1)} = \sum_{j=0}^{\infty}(-1)^j \frac{(i\beta)^j}{j!}(1-z)^j \tag{44}$$

Therefore the equation (41) can be written as follows:

$$I_c = \sum_{i=0}^{r}\sum_{k=0}^{i}\sum_{j=0}^{\infty} \frac{(i\beta)^j}{j!}(-1)^{(i+k+j)}\binom{r}{i}\binom{i}{k}\int_0^1 z^{r+i(\alpha-1)}(1-z)^{k\theta+j}dz$$

$$= \sum_{i=0}^{r}\sum_{k=0}^{i}\sum_{j=0}^{\infty} \frac{(i\beta)^j}{j!}(-1)^{(i+k+j)}\binom{r}{i}\binom{i}{k}B(r+i(\alpha-1)+1,k\theta+j+1) \tag{45}$$

where B is the beta function. Therefore the Chakravarty (1988) inequality measure is given by;

$$C_r = \left[\sum_{i=0}^{r}\sum_{k=0}^{i}\sum_{j=0}^{\infty} \frac{(i\beta)^j}{j!}(-1)^{(i+k+j)}\binom{r}{i}\binom{i}{k}B(r+i(\alpha-1)+1,k\theta+j+1)\right]^{1/r} \tag{46}$$

For empirical verification of the theoretical construct, we have to collect data.

Next, we will move to discuss about the collection, nature and source of data.

7. Data Collection

7.1. Introduction

To show the relative performance of the new functional form with compared to other functional forms, we have to do empirical analysis. For this empirical verification, we have to collect a set of data. Data have been obtained on the variables, monthly household income group in Taka, number of households corresponding to the income group, average monthly income per household in Taka corresponding to the income group. From our basic data we will derive the cumulative proportion of units which is represented by z and the cumulative proportion of income, which is represented by L(z). Suppose there are N households that are grouped into (T+1) income groups, $(x_0-x_1),(x_1-x_2),............,(x_T-x_{T+1})$.

Let n_i be the number of households corresponding to the ith income class. Then $f_i = \dfrac{n_i}{N}$ is the relative frequency of the ith income class. Let μ_i is the mean income of the ith income class. The proportion of households, earning income less than or equal to x_i is $z_i = \sum_{t=1}^{i} f_t$, where $\sum_{i=1}^{T+1} z_i = 1$, and the cumulative proportion of income received by this households is given by $(L(z))_i = \dfrac{\sum_{t=1}^{i} u_t f_t}{\mu}$, where $\mu = \sum_{t=1}^{T+1}\mu_t f_t$ and $\sum_{i=1}^{T+1}(L(z))_i = 1$. Then the Lorenz curve is the relationship between z and L(z).

Next we will discuss about the nature and source of data for the empirical analysis.

7.2. The Nature and Source of Data for Empirical Analysis

The success of any statistical and economical analysis ultimately depends on the availability of the appropriate data. It is therefore essential that we spend some times discussing the nature, sources and limitation of the data that may arise in empirical analysis. The difficulty lies in the availability and nature of the data. A particular problem facing the researcher is to obtain the appropriate data. In Bangladesh a reliable data is a golden deer. It is often difficult to obtain good reliable data with the necessary information required for a particular analysis. Missing value is a great problem in some data. Even in some experimentally collected data, errors of measurement arise from approximation and rounding off. Because of all these and many other problems the researchers should always keep in mind that the result of research may be affected by the quality of the data. Therefore, if in given situations the researchers find that the results of research are unsatisfactory the case may be not they used the wrong model but the quality of the data was poor. Some difficulties in finding suitable data are frequently encountered. The measurement of income presents particular difficulties, because there are a lot of wrong information.

For this study data were collected from BBS (Bangladesh Bureau of Statistics) publications " Household Expenditure Survey", corresponding to the years 1981-82, 1983-84, 1985-86, 1988-89, 1991-92, and 1995-96. On the basis of the given data set, next we will move for an empirical verification of the theoretical construct.

8. Empirical Analysis

To show the relative performance of the new functional form with compared to other functional forms, we have to estimate the residual sum of squares of these functional forms. For this, firstly, we estimated the parameter(s) values all of these functional forms by using the non-linear least squares method. And, then on the basis of these parameter(s) values we estimated the residual sum of squares of different functional forms. These estimated results are reported with the following tables:

Table (1)

Estimated New, Kakwani's and Podder's (1973, 76) Functional Forms

Year	New Function	Kakwani and Podder (1973)	Kakwani and Podder (1976)
1995-96	$L(z) = z^{0.1451} e^{0.4653(z-1)} (1-(1-z)^{0.5219})$	$L(z) = z e^{2.4811(z-1)}$	$L(z) = z^{0.3470} e^{3.7191(z-1)}$
1991-92	$L(z) = z^{0.1978} e^{0.4329(z-1)} (1-(1-z)^{0.6158})$	$L(z) = z e^{1.8337(z-1)}$	$L(z) = z^{0.2775} e^{2.7564(z-1)}$
1988-89	$L(z) = z^{0.1225} e^{0.4875(z-1)} (1-(1-z)^{0.5896})$	$L(z) = z e^{1.9401(z-1)}$	$L(z) = z^{0.1215} e^{3.0517(z-1)}$
1985-86	$L(z) = z^{0.2762} e^{0.3860(z-1)} (1-(1-z)^{0.5372})$	$L(z) = z e^{1.9141(z-1)}$	$L(z) = z^{0.1918} e^{2.9354(z-1)}$
1983-84	$L(z) = z^{0.2572} e^{0.1690(z-1)} (1-(1-z)^{0.6033})$	$L(z) = z e^{1.6327(z-1)}$	$L(z) = z^{0.3141} e^{2.5233(z-1)}$
1981-82	$L(z) = z^{0.2433} e^{0.2372(z-1)} (1-(1-z)^{0.5722})$	$L(z) = z e^{1.8621(z-1)}$	$L(z) = z^{0.2608} e^{2.7935(z-1)}$

Source: Data from BBS publications; own calculations

Table (2)

Estimated Kakwani's (1980), Rasche's et all (1980), Gupta's (1984) and Ortega's (1991) Functional Forms

Year	Kakwani (1980)	Rasche et all (1980)
1995-96	$L(z) = z - 0.8107 z^{1.0226}(1-z)^{0.4?}$	$L(z) = \left[(1-(1-z)^{0.5778}) \right]^{1/0.7274}$
1991-92	$L(z) = z - 0.7677 z^{1.0071}(1-z)^{0.56?}$	$L(z) = \left[(1-(1-z)^{0.6667}) \right]^{1/0.7075}$
1988-89	$L(z) = z - 0.7752 z^{1.0301}(1-z)^{0.54?}$	$L(z) = \left[(1-(1-z)^{0.6266}) \right]^{1/0.7294}$
1985-86	$L(z) = z - 0.6655 z^{0.9562}(1-z)^{0.4?}$	$L(z) = \left[(1-(1-z)^{0.5760}) \right]^{1/0.7964}$
1983-84	$L(z) = z - 0.6743 z^{0.9662}(1-z)^{0.5?}$	$L(z) = \left[(1-(1-z)^{0.6516}) \right]^{1/0.7596}$
1981-82	$L(z) = z - 0.7227 z^{0.9761}(1-z)^{0.50?}$	$L(z) = \left[(1-(1-z)^{0.6249}) \right]^{1/0.7437}$
	Gupta (1984)	**Ortega (1991)**
1995-96	$L(z) = z 11.9547^{(z-1)}$	$L(z) = z^{0.4536}(1-(1-z)^{0.5085})$
1991-92	$L(z) = z 6.2570^{(z-1)}$	$L(z) = z^{0.4802}(1-(1-z)^{0.5991})$
1988-89	$L(z) = z 6.9601^{(z-1)}$	$L(z) = z^{0.4288}(1-(1-z)^{0.5721})$
1985-86	$L(z) = z 6.7814^{(z-1)}$	$L(z) = z^{0.3104}(1-(1-z)^{0.5358})$
1983-84	$L(z) = z 5.1179^{(z-1)}$	$L(z) = z^{0.3699}(1-(1-z)^{0.5960})$
1981-82	$L(z) = z 6.4374^{(z-1)}$	$L(z) = z^{0.4053}(1-(1-z)^{0.5638})$

Source: Data from BBS publications; own calculations

Table (3)

Estimated Residual Sum of Squares of the New and Other Functional Forms

Year	New Function	Kakwani & Podder (1973)	Kakwani & Podder (1976)	Rasche et all (1980)	Kakwani (1980)	Gupta (1984)	Ortega (1991)
1995-96	0.00024	0.048461	0.0380	0.00047	0.00045	0.0485	0.00055
1991-92	0.00012	0.022296	0.0171	0.00054	0.00026	0.0223	0.00040
1988-89	0.00014	0.029831	0.0226	0.00095	0.0012	0.0298	0.00132
1985-86	0.000053	0.03901	0.0251	0.000055	0.000133	0.0319	0.000055
1983-84	0.000013	0.018593	0.0143	0.000035	0.000065	0.0186	0.000053
1981-82	0.000038	0.022466	0.0183	0.00021	0.00012	0.0225	0.00011

Source: Data from BBS publications; own calculations

From these estimated residuals it has been found that the residual sum of squares of the new functional form is smaller with compared to Kakwani's and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984), and Ortega's (1991) functional forms corresponding to different years. Therefore it can be concluded that our proposed functional form has demonstrated to give better fits for the Lorenz curve of a wide range of income distribution of Bangladesh with compared to these functional forms.

To test whether the parameters that are included in the new functional form are statistically significant or not in section 5, we specified various Lorenz curve hypotheses, and the value of F-test statistics under the maintained hypotheses H_0^1 - H_0^4, are estimated by using the non-linear least squares method. These estimated values are reported with the following table:

Table (4)

Estimated Values of F-test Statistics Under Various Lorenz Curve Hypotheses

Hypotheses	Estimated Values of F-Test Statistic					
	1995-96	1991-92	1988-89	1985-86	1983-84	1981-82
$H_0^1 : \alpha = 0$ p Value RSS	6.63735 0.02030 0.00034	22.6527 0.00021 0.00030	13.2305 0.00064 0.001015	67.5572 0.00000 0.00038	205.5125 0.000000 0.000285	65.57555 0.000000 0.000284
$H_0^2 : \beta = 0$ p Value RSS	20.3738 0.00035 0.00055	35.6337 0.00001 0.00041	5.11678 0.041475 0.001318	4.14282 0.03865 0.000055	29.95516 0.000272 0.000053	19.70379 0.001256 0.000111
$H_0^3 : \theta = 1$ p Value RSS	2480.35 0.00000 0.03797	2154.42 0.00000 0.01709	297.3429 0.00000 0.022589	5191.78 0.00000 0.025672	10797.74 0.000000 0.01432	4844.308 0.000000 0.018279
$H_0^4 : \alpha = 0, \beta = 0$ p Value RSS	512.151 0.00000 0.01582	1225.81 0.00000 0.01943	104.0067 0.00000 0.016087	675.6054 0.00000 0.65657	3847.753 0.000000 0.010212	1402.875 0.000000 0.010603

Source: Data from BBS publications; own calculations

From these estimated F-test statistics, it has been found that the null hypotheses , H_0^1, H_0^2, H_0^3 and H_0^4 are rejected at any reasonable significance level corresponding to different years. Therefore it can be concluded that all of the parameters that are included in the new functional form are very important for describing the Lorenz curve. And also we have estimated the residual sum of squares of these restrictive forms. We see that the estimated residual sum of squares of these restrictive forms are greater than the estimated residual sum of squares of the new functional form. Thus it can be concluded that the new functional form fits the data very well with compared these restrictive forms.

In section 6, we have derived the formulae of the Gini, Kakwani and Chakravarty inequality measures. Now, we also estimated these inequality measures on the basis of the new and also other functional forms.. We estimated the Kakwani and Chakravarty inequality measures for different values of "r" . These estimated values of these inequality measures are also reported with the following table:

Table (5)

Estimated Gini Index, and Kakwani and Chakravarty Inequality Measures for Different Values of "r"

On the Basis of the New and Other Functional Forms

Year	New Function	Kakwani & Podder (1973)	Kakwani & Podder (1976)	Rasche et all (1980)	Kakwani (1980)	Gupta (1984)	Ortega (1991)
Estimated Gini Index							
1995-96	0.4331	0.4916	0.4687	0.4359	0.4358	0.4916	0.4369
1991-92	0.3802	0.4091	0.3953	0.3843	0.3852	0.4091	0.38834
1988-89	0.3875	0.4241	0.4058	0.3963	0.3878	0.4242	0.3888
1985-86	0.3835	0.4205	0.4038	0.3912	0.3850	0.4205	0.3836
1983-84	0.3567	0.3787	0.3655	0.3568	0.3587	0.3787	0.3595
1981-82	0.3852	0.4131	0.3973	0.3877	0.3879	0.4131	0.3889
Estimated Kakwani Inequality Measure for r = 1.5							
1995-96	0.5030	0.5960	0.5555	0.5078	0.5063	0.5960	0.5102
1991-92	0.4514	0.5026	0.4742	0.4579	0.4519	0.5026	0.4581
1988-89	0.4563	0.5199	0.4834	0.4673	0.466	0.5199	0.4607
1985-86	0.4477	0.5158	0.4826	0.4551	0.4482	0.5158	0.4513
1983-84	0.4331	0.4674	0.4389	0.4233	0.4231	0.4673	0.4243
1981-82	0.4535	0.5073	0.4762	0.4571	0.4573	0.5073	0.4595
Estimated Kakwani Inequality Measure for r = 2							
1995-96	0.5516	0.6637	0.6057	0.5555	0.5521	0.6638	0.5596
1991-92	0.4996	0.5655	0.5219	0.5085	0.5001	0.5656	0.5101
1988-89	0.5019	0.5840	0.5285	0.5155	0.5020	0.5841	0.5103
1985-86	0.4909	0.5796	0.5293	0.4975	0.4913	0.5796	0.4953
1983-84	0.4684	0.5278	0.4831	0.4689	0.4683	0.5278	0.4711
1981-82	0.4996	0.5706	0.5238	0.5041	0.5040	0.5706	0.5081
Estimated Kakwani Inequality Measure for r = 2.5							
1995-96	0.5842	0.7103	0.6350	0.5902	0.5845	0.6503	0.5959
1991-92	0.5346	0.6103	0.5515	0.5462	0.5348	0.6103	0.5492
1988-89	0.5346	0.6293	0.5549	0.5511	0.5342	0.6294	0.5474
1985-86	0.5228	0.6248	0.5575	0.5283	0.5226	0.6248	0.5278
1983-84	0.5020	0.5711	0.5103	0.5029	0.5015	0.5711	0.5062
1981-82	0.5332	0.6155	0.5532	0.5388	0.5379	0.6155	0.5442
Estimated Chakravarty Inequality Measure for r = 1.5							
1995-96	0.4537	0.5204	0.5024	0.4557	0.4569	0.5204	0.4561
1991-92	0.3976	0.4323	0.4226	0.4009	0.3981	0.4323	0.3996
1988-89	0.4057	0.4484	0.4358	0.4137	0.4064	0.4484	0.4099
1985-86	0.4004	0.4445	0.4326	0.4004	0.4011	0.4445	0.4030
1983-84	0.3722	0.4001	0.3914	0.3723	0.3726	0.4001	0.3745
1981-82	0.4024	0.4367	0.4249	0.4048	0.4054	0.4367	0.4054
Estimated Chakravarty Inequality Measure for r = 2							
1995-96	0.4696	0.5426	0.5267	0.4710	0.4730	0.5427	0.4708
1991-92	0.41103	0.4502	0.4426	0.4138	0.4117	0.4502	0.4119
1988-89	0.4198	0.4669	0.4570	0.4271	0.4208	0.4670	0.4228
1985-86	0.4134	0.4630	0.4534	0.4137	0.4143	0.4629	0.4160
1983-84	0.3843	0.4165	0.4101	0.3841	0.3848	0.4164	0.3861
1981-82	0.4157	0.4547	0.4451	0.4179	0.4189	0.4547	0.4181
Estimated Chakravarty Inequality Measure for r = 2.5							
1995-96	0.4824	0.5601	0.5455	0.4832	0.4859	0.5601	0.4826
1991-92	0.4218	0.4666	0.4583	0.4241	0.4226	0.4644	0.4219
1988-89	0.4311	0.4818	0.4736	0.4379	0.4322	0.4819	0.4331
1985-86	0.4238	0.4776	0.4696	0.4244	0.4248	0.4777	0.4264
1983-84	0.3939	0.4295	0.4248	0.3936	0.3946	0.4295	0.3954
1981-82	0.4264	0.4691	0.4609	0.4285	0.4297	0.4692	0.4284

Source: Data from BBS publications, own calculations

From these different estimated inequality measures it has been found that the new functional form gives the lowest level of inequality with compared to Kakwani's and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984) and Ortega's (1991) functional forms. Therefore, it can be concluded that the Kakwani's and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984) and Ortega's (1991) functional forms are so restrictive for describing Lorenz curves and probably lead to an upward bias of estimated inequalities. Also from these estimated inequality measures it can be concluded that, in 1995-96, inequality is higher with compared to the previous years. This indicates that the income distribution of Bangladesh is going to be worsened and may be continued of worsening to the years ahead. This indicates that in Bangladesh, the unemployment rate is increasing and the social welfare policy of Bangladesh is going to be worsened. From the estimated values of the Kakwani and Chakravarty inequality measures it has been found that for increasing value of "r", the estimated values of these measures will be increased. This indicates the sensitivity of these inequality measures to change in "r". This bears a significant importance for decision making about the value of "r" to be attached by the society. From our analysis, it has been found that, the Kakwani and Chakravarty measures are highly correlated with the Gini index. The correlation coefficient between Kakwani and Gini index is about 0.9996 and between Chakravarty and Gini index is about 0.8953. So, Kakwani measure is highly correlated with the Gini index than Chakravarty measure.

Next we will move for an overall discussion and conclusion.

9. Discussion and Conclusion

In the past decades, many authors have proposed different functional forms for estimating Lorenz curves from grouped data. Most of these functional forms do not fit the data very well. That is why, in this paper, we proposed a new functional form for estimating Lorenz curves, based upon the studies of the income distribution of Bangladesh, which fits the data very well with compared to Kakwani's and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984) and Ortega's (1991) functional forms.

For empirical verification of the theoretical construct a set of data has been collected from BBS (Bangladesh Bureau of Statistics) publications "Household Expenditure Survey" corresponding to different years. On basing upon these data, we estimated the parameter(s) values of different functional forms by using the non-linear least squares method, which are reported in tables (1) and (2). Now, on the basis of these estimated parameter(s) values of different functional forms, we estimated the residual sum of squares of the different functional forms, which are reported in table (3). From these estimated residuals, it has been found that the residual sum of squares of the new functional form is smaller with compared to other functional forms. Thus, it can be concluded, the new functional form has demonstrated to give better fits for the Lorenz curve with compared to Kakwani`s and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984), and Ortega's (1991) functional forms for a wide range of income distribution of Bangladesh.

In order to test whether the parameters that are included in the new functional form are statistically significant or not, we estimated the F-test statistics by using the non-linear least squares method under various Lorenz curve hypotheses. These estimated values are reported in table (4). From these estimated results it has been found that all of the parameters that are included in the new functional form are most important for describing the Lorenz curve.

Also, on the basis of the new functional form we derived the formulae of the Gini, Kakwani and Chakravarty inequality measures. We also estimated these inequality measures on the basis of the new and other functional forms. We

estimated the Kakwani and Chakravarty inequality measures for different values of "r". These estimated results are also reported in table (5). From these estimated inequality measures it can be concluded that the new functional form gives the lowest level of inequality with compared to Kakwani's and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984) and Ortega's (1991) functional forms. Therefore, it can be concluded that the Kakwani's and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984) and Ortega's (1991) functional forms are so restrictive for describing Lorenz Curves and probably lead to an upward bias of estimated inequality.

Also from these estimated inequality measures it can be concluded that in 1995-96, the dispersion of the household income is higher than to the previous years in Bangladesh. This indicates that in Bangladesh, the income distribution is going to be worsened and may be continued of worsening to the years ahead. The major cause for such an increase may be explained by the fact increase in unemployment rate specially increased in unskilled workers, and the social welfare policy of Bangladesh is going to be worsened from year to year. Also, population growth is another factor for increasing inequality, which reinforces the polarization process by increasing pressure on land, causing increasing the unemployment rate and widening the gap between the poor and the rich. Inflationary pressure is also another important factor that accentuates inequality because this hits the low income groups severely, lowering their real income, as only basic necessities for major share of their total expenditure. Natural disaster like as flood is also another factor of increasing the inequality.

From the estimated results of Kakwani and Chakravarty inequality measures it has been found that for increasing value of "r", the estimated values of these inequality measures are increased. This indicates the sensitivity of these inequality measures to change in "r". This bears a significant importance for decision making about the value of "r" to be attached by the society. Also, from our analysis it has been found that the Kakwani and Chakravarty inequality measures are highly correlated with the Gini index. But Kakwani measure is highly correlated with Gini index with compared to the Chakravarty measure. The correlation coefficient between Kakwani and Gini index is about 0.9996 and between Chakravarty and Gini index is about 0.8953.

So our analysis bears a number of important policy implications. The strategy for economic development should aim to increase per capita income, expansion of productive employment and greater equality in the distribution of benefits accrued from the development and welfare policies. This implies decreasing inequality by increasing productivity of the target groups which includes landless laborers, small farmers, share croppers, artisans etc. and provides them with greater access to goods and services. Reconstruction and resurgence of scarce resources together with redistribution of public services, are therefore, considered necessary for the reduction in income inequality in Bangladesh.

Finally, it can be concluded that our proposed functional form is better with compared to Kakwani's and Podder's (1973, 76), Rasche's (1980), Kakwani's (1980), Gupta's (1984) and Ortega's (1991) functional forms for estimating Lorenz curves from grouped data. 📖

References

1. Chakravarty, S.R., 1988, "Extended Gini Indices of Inequality", *International Economic Review,* 29, 147-156.
2. Gastwirth, J. L., 1971, "A General Definition of the Lorenz Curve", *Econometrica,* 39, 1037-1039.
3. Gupta, M.R., 1984, "Functional Form for Estimating the Lorenz Curve", *Econometrica*, 52, 1313-1314
4. Kakwani N., 1980a, "Income Inequality and Poverty", *Oxford*: Oxford University Press.
5. Kakwani N., 1980b, "On a Class of Poverty Measures", *Econometrica,* 48, 437-446
6. Kakwani N., and Podder, N., (1973), "On the Estimation of Lorenz Curves from Grouped Observations", *International Economic Review*, 14-2, 278-291
7. Kakwani N., and Podder, N., 1976, "Efficient Estimation of Lorenz Curve and Associated Inequality Measures from Grouped Observations", *Econometrica,* 14-2, 278-291
8. Lorenz, M. O., 1905, " Methods of Measuring the Concentration of Wealth", *Journal of American Statistical Association*, 9, 209-219
9. Rasche, R. H.,Gaffney. J. Koo, A. Y., and Obst, N., 1980, "Functional Forms for Estimating the Lorenz Curve", *Econometrica,* 48, 1061-1062

Notes