PARAMETRIC MODELING OF HOUSEHOLD INCOME DISTRIBUTION IN THE PUNJAB, PAKISTAN

PARAMETRIC MODELING OF HOUSEHOLD INCOME DISTRIBUTION IN THE PUNJAB, PAKISTAN

Muhammad Shakeel[1], Ijaz Hussain[2], Mian Muhammad Arif [3], Muhammad Ameen[4] and Muhammad Ahsan ul haq[5]

[1]Department of Statistics, Govt Degree College Boys, Hujra Shah Muqeem, Okara-Pakistan

[2]Department of Statistics, Quaid-i-Azam University, Islamabad-Pakistan

[3]University of Education, Township Campus, Lahore-Pakistan

[4]Punjab Bureau of Statistics, Lahore-Pakistan

[5]College of Statistical and Actuarial Sciences Punjab University Lahore-Pakistan

*Corresponding Author: shakeel_shera@yahoo.com

ABSTRACT: Accurate and precise measurement of income inequality is a crucial problem for policy makers to formulate effective strategies for social welfare. Since the correct specification of the income density model has a greater importance and is considered the basis of all the parametric inequality measures. Generalized Beta distribution of second kind (GB2) is considered in order to model the income distribution of Punjab province for the year 2004 and 2008. It is observed that the GB2 models fits reasonable on our both the data sets (year 2004, and 2008) better than nested alternatives models (BII, Dagum and Singh). Performance of the fit is evaluated through graphically and numerically measures. Related parametric measures of income inequality like Gini index, generalized entropy measure, two percentile ratios and Lorenz curve illustrate that income inequality is increased in the province of Punjab during the years 2004-2008.

***Key words:** Income Inequality; GB2; Gini Index; Generalized entropy measure; Lorenz Curve*

INTRODUCTION

Even from everyday experience, one can easily understand that almost without any exception income and wealth in a society are unequally distributed among its people and from ancient times, this inequality has been a constant source of irritation in all the societies and countries across the world. There are several non-trivial issues and questions related to this observation. In fact, the issue of inequality in terms of income and wealth has been perhaps the most fiercely debated one in economics [1].

On a global and regional level, monitoring and measuring inequality tells us whether wealth is becoming more concentrated or there is decreasing world or regional income inequality. Information from income inequality measures are employed to measure welfare, poverty and inequality, to assess changes comparatively in these measures over time and across countries. These inequality measures are also used to check the effectiveness of some newly implemented polices regarding some social welfare and taxation programs [2]. The estimations of income inequality plays a vital role in decision making in economic policies and various fields of social politics and well-being of society. Typical and popular parametric inequality measures are coefficient of variation, classical Gini coefficient and Lorenz curve which are based on proportion of population below a specified threshold or the expected value of a function over that part of the income distribution below a specified threshold [3]. Estimates of these parametric inequality measures and Lorenz curve totally depend on the income distributions parameters and the method of parameters estimation [2]. Thus the appropriate income size distributions models have a central importance in assessing real estimates of income inequality measures.

The earliest model of measuring income distribution and measurement of income inequality has been introduced through one parameter Pareto distribution by Vilfredo Pareto [4]. However, Pareto distribution usually possesses better fit for upper tail but it is not useful for fitting the entire range of income data adequately [5]. After the one parameter Pareto distribution model, the two parameters lognormal distribution [6] has been proposed but this lognormal distribution fits well on middle part of income ranges but gives poor fit at upper tail. Later on the other two parameters models such as the Fisk [7], Gamma [8] and Weibull [9] distributions have been suggested and employed in the income size distributional literature. In the mid-1970, the three parameters models such as the singh-maddala [10], dagum [11] and generalized gamma [12] distributions have been introduced which include Pareto, Fisk, log-normal, gamma and weibull distributions as a special cases. In 1984, MacDonald [13] presented the four parameters income distribution models which are known as generalized beta of the first (GBI) and second kind (GB2). The GBI and GB2 models encompass and incorporate all the previously mentioned income distribution models as special, nested or limiting cases.

Numerous other frequently employed distributions (see [13]) including exponential, Pareto, Weibull, Dagum type I, generalized beta of first and second type have been used to model the personal and household income by different researchers on empirical income datasets. MacDonald [13] has fitted the above models to the US family income data sets of 1970, 1975 and 1980 and concluded that GB2 distribution provided the best relative fit and the Singh-Maddala distribution provided the better fit than the GBI distribution model. In 2002, [14] have compared the performance over time and across countries of GB2 distributions and its various nesting alternative models for income distributions. They have shown that the GB2, Dagum and Weibull are best fitting models among the four, three and two parameters considered income distributions models respectively. Brzezinski [15] has found that GB2 model has also been fitted better than its nested alternative models (Dagum and Singh-Maddala distributions) to the income distribution of Poland and Hungary. However for the Slovak Republic and Czech Republic the Dagum model performed better fit as compared to GB2 model for the year of 1990 and 2000. [16] measured the various income inequality measures including Gini index of ten countries of South and Southeast Asian region including Pakistan using GB2 and its three popular special cases such as B2, Dagum and Singh-Maddala distributions for the years 1992, 2000, 2005 and 2008 but the weakness of their study is that they used grouped data which are available in the form of population shares and corresponding expenditure shares for a number of expenditure classes rather than use of raw data.

For a successful modeling of income and wages, a flexible and positively skewed distribution with long right tail and high density at lowest percentile is necessary. There is substantial and wide literature describing the properties, inequality indices, estimation procedures, and applications of GB2 family of distributions see [13,17,18,19], with further information on inequality measures provided by [2,20,21,22,23,24].

In present paper we are focused on a prevailing and more refined beta-type family of income distribution, Generalized beta II (GB2) distribution and its popular special and limiting cases like exponential, Gamma, Weibull, Log-normal, Fisk, Singh-Maddala, Dagum and Beta II that have been repeatedly used and are considered successful in describing empirical personal and household income data sets and have also been employed for calculation of parametric inequality measurement indices like Gini index, percentile ratios and Lorenz curve. Despite wide spread use of GB2 as an income distribution, it is popular, flexible and has been widely acknowledged to be an adequate and well suited income family distribution model, providing satisfactory goodness of fit with relative parsimony and subsuming many other income distributions models as a special, nested or limiting cases. Furthermore our both data set do not show any evidence of a possible use of mixture distribution. For this reason, in order to find the best fit for the distribution of household income, the possible use of GB2 and its nested alternative distributions model is investigated and their relative performance have been evaluated through the Bayesian information criterion (BIC), Chi-square goodness of fit with supplemented others measures like sum of square of error (SSE) and sum of absolute error (SAE) measures. Likelihood Ratio test is also employed for selecting best fitting income distribution on our both household income datasets. The main aim for this empirical analysis is to find or investigate the best fitted income size distribution from above cited distributions and also to compare the level of increasing disparity or inequality of income through parametric inequality measures of above mentioned income distribution models between the years 2004 to 2008 in the household income sphere in the Punjab province Pakistan.

MATERIALS AND METHODS

DATA DESCRIPTION AND RESOURCES

We use individual household per capita income data by combining primary and secondary sources of income. The two datasets coming from Multiple Indicator Cluster Survey (MICS) of Punjab Bureau of Statistics Planning and Development Department, Government of the Punjab, Pakistan. For more detail on MICS, visit http://www.bos.gop.pk. In present paper, two trimmed samples of household income expressed in nominal local currency unit between the ranges of 500 to 90,000 rupees from MICS (2003-04) and (MICS 2007-8) are selected. Descriptive statistics for income variable of our both the data sets are presented in Table 1 and histograms are shown in Figure 1.

Table 1. Descriptive Statistics of per capita Household income and Histogram

Descriptive Sample Statistics	2003-04	2007-08
Count	29630	87766
Average	7498.64	10815.19
Standard Deviation	7742.45	10778.1
Coefficient of Variation	103.2514	99.6571
Minimum	508	516
Maximum	89167	89667
Range	88659	89151
Skewness	4.189834	2.88
Kurtosis	29.309	14.15

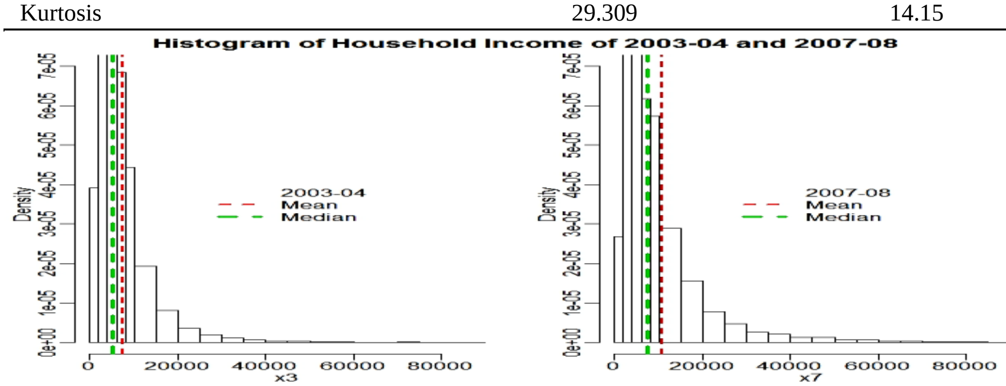

Figure 1. Histogram of per capita household income

GENERALIZED BETA II DISTRIBUTION FAMILY

Generalized Beta distribution of the second kind (GB2) which is a very popular and flexible four parametric income distribution model. It has been introduced by McDonald [13] and shows that many of the previously mentioned distributions can be represented as limiting or special cases of GB2. It is known as empirically best fitted distribution.

$$f(x) = \frac{\alpha x^{\alpha p-1}}{b^{ap}B(p,q)[1+(x/b)^{\alpha}]^{p+q}}$$
$$where\ \ x,a,b,p,q > 0$$

Here $B(u,v) = \Gamma(u)\Gamma(v)/\Gamma(u+v)$ is beta function and $\Gamma(.)$ is the gamma function, all four parameters are positive where b the scale parameter, and all other are shape parameters. In the shape parameters, a represent the overall shape, q governs right tail and p the left tail. The cumulative distribution function (cdf) of GB2 distribution involves an infinite series so it does not have an explicit form.

The GB2 distribution can be expressed as a mixture of inverse generalized gamma and generalized gamma distribution [21]. Using this parameterization, it is very easy to see relationship within the GB2 family of distributions. The three parameters BII, Dagum and Singh-Maddala are special cases of GB2 distribution

$$BII(x;b,p,q) = GB2(x;a=1,b,p,q)$$
$$Dagum\ (x;a,b,p) = GB2(x;a,b,p,q=1)$$
$$SM(x;a,b,q) = GB2(x;a,b,p=1,q)$$

The k[th] order moments of GB2 (existing only for $-ap < k < aq$)is defined as

$$E(X^k) = \frac{b^k B(p+\frac{k}{a}, q-\frac{k}{a})}{B(p,q)}$$

A convenient way to visualize the relationship of GB2 with other special or nested income size distribution models is presented in Figure 2.

The higher the distribution is on a branch of the hierarchy tree in the above Figure 1, the better it would perform as measured by the same criterion [13]. Thus, the GB2 model should provide at least as good fit as any special or nested distribution model. However a limiting or special case might equal the GB2 model.

MEASURES OF INCOME INEQUALITY

The cumulative distribution function of income uniquely characterizes distributional characteristics of income; alternative measures can facilitate a comparison of the relative inequality of two distributions of income. For example “the Lorenz curve depicts the relationship between the percent of income received by different percentages of a given population”. Lorenz curve is defined with reference to a given distribution function.

$$L(u) = \frac{1}{\mu}\int_0^u F^{-1}(t)dt$$

Lorenz curve is entirely contained into a square, because the upper limit of the integral u is defined over the interval [0, 1] and $L(u)$ denotes the fraction of total income that holders of the lowest uth fraction of the income possess and μ is the mean income [21]. Lorenz curve is continuous function if the underlying variable has positive values and has the density. It is always situated below the 450 line or equal to it. In comparing two populations using Lorenz curve, population 2 is said to be more egalitarian than population 1 if $L_1(u) \le L_2(u)$ for all $u, 0 < u < 1$.

Numerous scalar measures of inequality have been considered and used in the literature, including Gini coefficient, coefficient of variation and a family of generalized entropy inequality measures $GE(\gamma)$. The classical Gini Coefficient can be interpreted as twice the area of concentration between the Lorenz curve and 45 degree line of perfect equality. Its popularity contains or stems from its simplicity both in interpretation as well as computation. The range of Gini value is between zero and one. Zero value of Gini shows the perfect equality that every person has equal or identical income, and one value shows perfect inequality that one person has all the income.

The Gini index of income inequality is most sensitive to the income difference around the mode of the distribution and

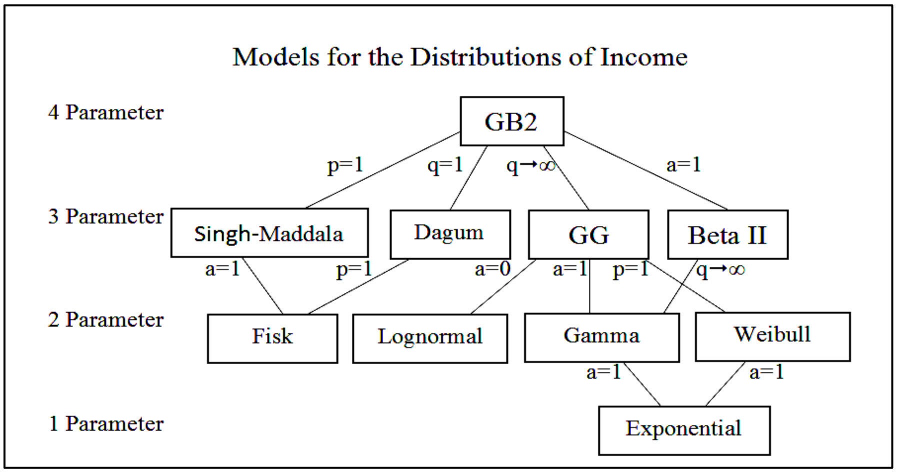

Figure 2. Beta-type distribution tree.

therefore, it is not appropriate and suitable to detect distributional changes that occur in the higher end or in the lower end of the distribution. To overcome this difficulty, a family of distribution sensitive generalized entropy measures $GE(\gamma)$ has been presented [25]. The smaller the values of parameter γ is, the more sensitive is $GE(\gamma)$ to the income difference at the lower tail of the distribution, the higher values of it is, the more sensitive is $GE(\gamma)$ to the income difference at the upper tail of the distribution. The most prominent members of $GE(\gamma)$ family include the Theil index, $GE(1)$and half the square of coefficient of variation GE(2) [15]. In this paper, $GE(2)$ inequality measure will be used because it has been shown that this inequality measure is particularly sensitive to the presence of extremely large income observations. The value of GE measures is between the ranges of 0 to ∞ where zero means the perfect equality while higher values of GE represent higher levels of inequality.

In the GB2 model and its nested models, the relationship between model parameters and inequality indices is very complex. The analytical expressions or equations for Gini index, $GE(2)$ and Lorenz curve in terms of distributional parameters of GB2 distribution models and all its nested models employed for this study are given in Table 2 (given in Appendix). All these formulas and expressions can be found in literature such as [13] [19] [22] [24] [26].

DATA ANALYSIS AND DISCUSSION

All income size distribution models have been fitted on individual household income datasets using maximum likelihood estimation (MLE) and the inequality measures are calculated directly from the MLE of these parameters estimates. For fitting models to data, [28] and GB2 [29] packages of R software are used. MLE is probably the most common method of estimating the parameters of the above mentioned income size distribution models. However evidences show that the MLE out performs the individual or ungroup income data which is also used here in this empirical study.

In order to select best fitted distribution of the nine competitive income size distributions regarding relative fitting accuracy and performance, both graphical and numerical methods are employed. Visually and graphically, the most popular method is the quantile-quantile (q-q) plot which plots the theoretical versus sample quantile of the variable. If the estimated distribution model fits the data perfectly, the resulting q-q plot would definitely coincide with the 45-degree line. Numerically, we evaluated sum of square of error (SSE), sum of absolute error (SAE), Chi-square goodness of fit in addition to likelihood ratio test to estimate income size distribution models. A brief introduction of these numeric measures is given below.

BAYESIAN INFORMATION CRITERION

The Bayesian Information Criterion (BIC) is employed for identification of an optimum distribution model in a class of competing distributions models.

$$BIC = k\ln(n) - 2ln(L(\hat{\theta}))$$

Where $L(\hat{\theta})$ likelihood function, k and n are number of estimated parameters of the fitted distribution model and number of observations in the data set respectively.

LIKELIHOOD RATIO (LR) TEST

In order to compare GB2 models and its three parameters nested model we have employed the likelihood ratio test.

$$LR = 2(\hat{l}_u - \hat{l}_r) \sim \chi^2(h)$$

Where $\hat{l}_u$ is the log-likelihood function of the unconstrained (GB2) and $\hat{l}_r$ is the log-likelihood function of restricted model (BII, Singh-Maddala and Dagum etc.) and h is the difference in the number of parameters in both compared models.

CHI-SQUARE, SSE AND SAE

Chi-square, SSE and SAE are other relative measures for comparing the goodness of fit performance of estimated distributions models.

$$SSE = \sum_{i=1}^{k=20} \left(\frac{n_i}{N} - P_i(\hat{\theta})\right)^2$$

$$SAE = \sum_{i=1}^{k=20} \left|\frac{n_i}{N} - P_i(\hat{\theta})\right|$$

$$\chi^2 = N \sum_{i=1}^{k=20} \left[\left(\frac{n_i}{N} - P_i(\hat{\theta})\right)^2 \Big/ P_i(\hat{\theta})\right]$$

$\hat{\theta}$ is a vector of estimated parameters, N is total number of observed counts in the data sets, the values of k is number of grouping classes. In this study $k = 20$ i-e, breaks are 20 between the range of 500 to 90,000 with equal class interval of 4500.

EMPIRICAL RESULTS

The main objective of this section is to investigate the possible use of GB2 distribution and its nested alternative distribution models in describing the household per capita income distribution in Punjab, the province of Pakistan. Maximum likelihood estimates are employed to estimate the unknown population parameters for each probability distribution models and results are presented in Table 3.

The smaller values of the standard error indicate that all the parameters are very precisely estimated. Each distribution has been fitted to both of our data sets and goodness of fit criteria is calculated including log-likelihood value as well as chi-square, SSE and SAE values.

The Fisk and Dagum distributions are clearly the best fitting two and three parameter models respectively for the year 2004. This holds true for Fisk distribution regarding log-likelihood, BIC and chi-square criteria while lognormal distribution is showing slight improvement on the basis of SAE and SSE criteria. This also demonstrates the advantages of the three parameters Dagum distribution over the two parameters gamma, Weibull, lognormal and Fisk distribution models. GB2 distribution model provides very marginal improvement regarding log-likelihood criterion, but it is defeated by very short margin to the three parameters Dagum model especially on the basis of chi-square criterion.

Table 3. Results of Maximum likelihood estimates of selected distribution models for Household Income

Distribution	**Parameters estimates for 2003-04**	**Parameters estimates for 2007-08**
Exponential	λ =0.00013 (1e-06)	λ = 9e-05 (3e-07)
Gamma	α=1.784 (0.0135), γ= 0.00024 (0.000002)	α=1.58 (0.007), γ= 0.00015 (1.0e-06)
Weibull	α=1.2270 (0.0056), β=8108.232 (40.38)	α=1.194 (0.0031) , β=11583.19(34.46)
Lognormal	μ=8.62 (0.0043), σ =0.748 (0.0031)	μ=8.94 (0.0028), σ = 0.8224 (0.002)
Fisk	a=2.396(0.0116), b=5414.41(22.74)	a=2.14 (0.006), b=7547.132 (20.64)
Singh-Maddla	α=2.71 (0.0255), q= 0.736 (0.0141) β=4509.96 (51.54)	α=2.25 (0.0119), q=0.873 (0.0103) β=6863.4 (56.97)
Dagum	α=2.133 (0.0178) , p=1.484 (0.0375) β=4175.56 (75.045)	α=1.945 (0.0095) , p=1.365 (0.02) β=5998.034(66.359)
Beta II	b=2682.37 (136.18), p= 5.97 (0.1871) q=3.144 (0.0537)	b=5182.294 (133.664), p=4.11 (0.0575) q=2.94 (0.0301)
Generalized Beta II	a=1.851 (0.0792), b=4039.936 (99.037) p=1.86032 (0.1384), q=1.231(0.0767)	a=0.549(0.0511), b=2753.55 (533.28) p=13.98 (2.94), q=8.191 (1.311)

Note: Standard errors of estimates are given in parentheses.

Table 4. Goodness of Fit of GB2's Nested Distributions for the year 2004 & 2008

Distribution	**Household Income Goodness of fit 2003-04**					**Household Income Goodness of fit 2007-08**				
	Loglik	BIC	χ^2	SSE	SAE	Log-Lik	BIC	χ^2	SSE	SAE
Exponential	-294003	588016.3	8458.92	0.0105	0.2159	-902998.6	1806009	5924.3	0.00595	0.16177
Gamma	-291499	583018.6	7536.76	0.0124	0.2072	-898075.5	1796174	19732.4	0.00619	0.19304
Weibull	-292835.9	585692.4	8458.92	0.0125	0.2306	-900537.7	1801098	1689	0.00601	0.19408
Lognormal	-288753.6	577527.8	1042.45	0.0019	0.0886	-892099.9	1784223	987.9	0.00119	0.07217
Fisk	-288591.5	577203.6	346.81	0.0023	0.0894	-892666.4	1785356	1446.8	0.00363	0.10810
Singh-Maddla	-288485.9	577002.7	129.85	0.0011	0.0565	-892616.9	1785268	1396.3	0.00381	0.10848
Dagum	-288452.3	576935.5	103.15	0.0008	0.0506	-892431.7	1784898	1114.8	0.00310	0.09878
Beta II	-288501.6	577034.1	146.62	0.0007	0.0507	-892038.4	1784111	739.8	0.00166	0.07041
GBII	-288447.3	576935.8	112.28	0.0008	0.0509	-891998.2	1784042	652.3	0.00116	0.06454

On the other hand, the lognormal, BII and GB2 are clearly the best fitting two, three and four parameters models respectively for the year 2008 regardless of the criterion used for the comparison. This can be seen by visual inspection of density plots overlaid with the fitted pdf in Figure 3 and q-q plot in Figure 4 confirm that all the distribution models numeric goodness of fit measures defined positions remain v alid and equivalent to observational and visual criteria.

In Figure 4, it can be easily observed in the q-q plots that GB2 distribution model gives the best fit for both the data sets. Other models are visible worse, especially for higher quintiles. It can also be observed that the two parameter models show a significantly worse fit than the three parameters models. The nested relationship of the distribution models guarantees that the GB2 will fit the data at least as well their special cases. Similar results are obtained by additionally performing the likelihood ratio tests presented in table 5, suggest that the GB2 model is preferred to Singh-Maddla, Dagum and BII distribution models for both the datasets under study.

For inequality measures of goodness of fit is evaluated numerically in table 6 by comparing the sample values of chosen distributional indicators with their counterparts implied by the fitted distributional models. The results suggest that for most of the inequality indices, the best fitting models produce indices values that are often in a close agreement with the corresponding sample values. The one exception is the top-sensitive inequality index GE(2) which differ for the first and second data set for the best fitting GB2 distribution model by about as much as 166% and 20% respectively by its sample counterpart value. GE(2) is also differing by its sample counterpart for the best fitting Dagum model for the first data set and BII model for the second data

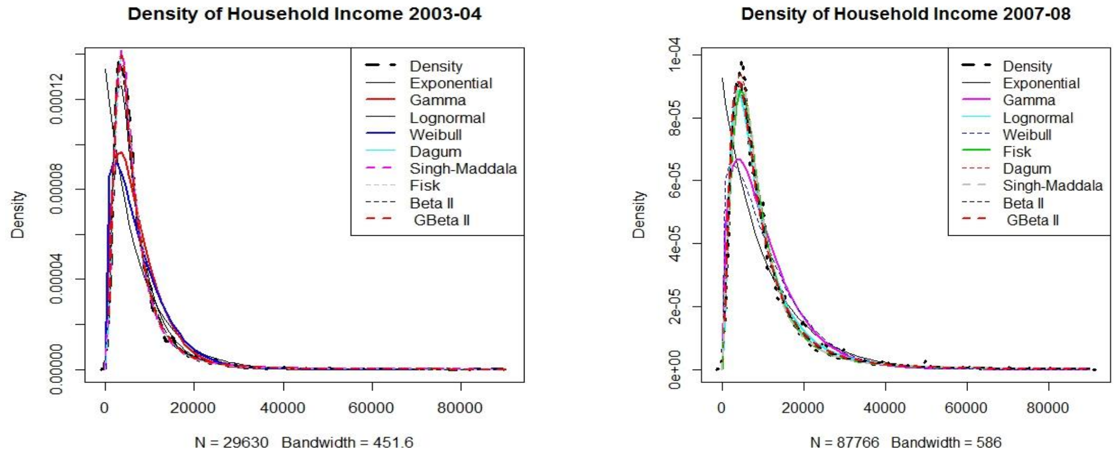

Figure 3: Distribution of household income for 2004 and 2008

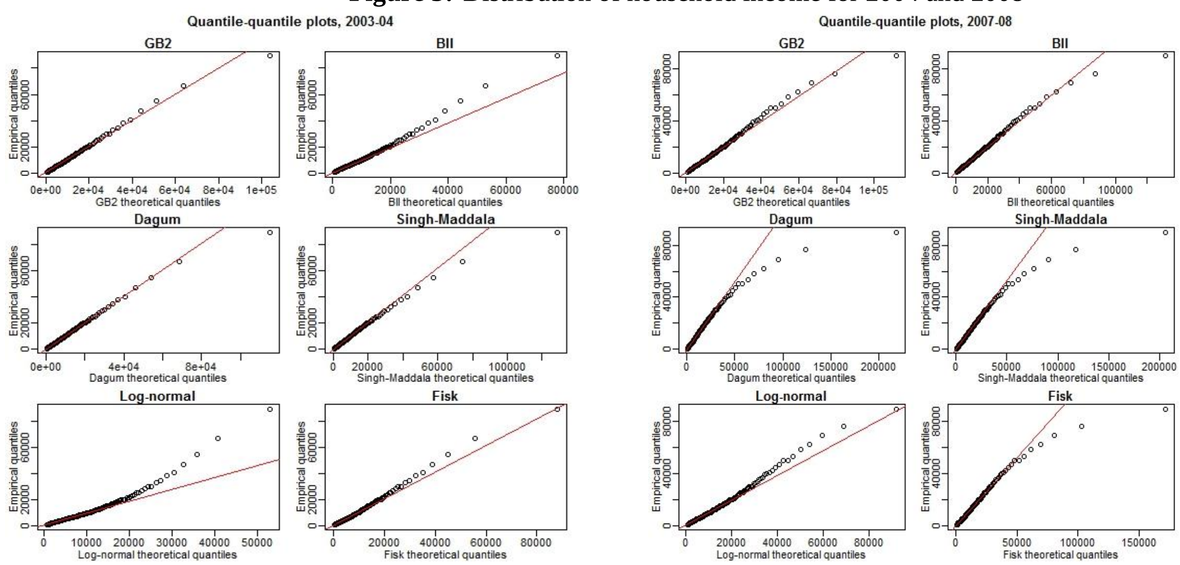

Figure 4: Quantile-Quantile plots for the selected statistical distributions models

Table 5. Likelihood Ratio test for model selection

Year	BII vs GB2		Singh-Maddala vs GB2		Dagum vs GB2	
	LR	P value	LR	P value	LR	P value
2003-04	108.56	0.0000	77.07	0.0000	9.95	0.0016
2007-08	80.3245	0.0000	1237.33	0.0000	867.10	0.0000

set in three parameters models about as much as 413% and 56% respectively. These facts reflect the high sensitivity of some inequality indices to the presence of extremely large incomes. It is also noted that the estimates implied by fitted parametric models seem to be much less sensitive to extreme observations than sample estimates. On the other hand it is also worth stressing here that both types of estimate (the sample estimates and estimates implied by the fitted model) for the inequality measures percentile share ratios and Gini coefficient employed in our analysis differ by no more than 3.5% and 2.4% respectively for the best fitting GB2 model for both the data sets. This suggests that the GB2 distribution model is quite successful in describing the inequality of income distribution in the Punjab province during the year 2004 to 2008, at least if one is focusing on Gini coefficient. In table 6 sample estimates of all the four widely used inequality indices are presented, the GE(2) index, the Gini index and two percentile ratios. According to Gini index and two percentile ratios estimates suggest that the inequality increased but the GE(2) suggest otherwise.

The parameters estimates presented in table 3 have also been used to build estimated Lorenz curve by applying parametric equations given in Table 1. The curves for both the years are presented in Figure 4 together with empirical Lorenz curve estimates. GB2, and Dagum for the year 2004, and GB2, and BII for the year 2008 lead to estimated Lorenz curve exhibiting a degree of inequality that is much more in line with the empirical Lorenz curves for the respective years.

Table 6: Distributional and inequality measures summary statistics of household income

Distributions	Year	Mean	Median	P90/P10	P75/P25	GE(2)	Gini	S.D
Empirical	2003-04	7498.6	5250.0	6.5	2.6	0.53	0.42	7742.5
Exponential		7498.6 (0%)	5197.7 (-1%)	21.9 (235.6%)	4.8 (88.3%)	0.50 (-5.7%)	0.50 (19%)	7692.3 (-0.64%)
Gamma		7498.6 (0%)	6153.4 (17.2%)	8.42 (29.3%)	3.0 (17.6%)	0.89 (67.9%)	0.39 (-7.1%)	5565.3 (-28.1%)
Weibull		7584.8 (1.1%)	6014.6 (14.2%)	12.4 (89.7%)	3.6 (40.6%)	0.34 (-35.8%)	0.43 (2.4%)	6214.5 (-19.7%)
Lognormal		8028.2 (7.1%)	5523.2 (5.2%)	6.8 (4.5%)	2.7 (7%)	0.37 (-30.2%)	0.40 (-4.8%)	5881.1 (-24%)
Fisk		7345.85 (-2%)	5414.4 (3.1%)	6.3 (-3.8%)	2.5 (-2.3%)	0.94 (76.7)	0.42 (0%)	10048.5 (29.8%)
Dagum		7725.104 (3%)	5324.2 (1.4%)	6.3 (-3.4%)	2.5 (-1.2%)	2.72 (413.2%)	0.45 (7.1%)	18032.2 (132.9%)
Singh-Maddla		7815.817 (4.2%)	5319.6 (1.3%)	6.2 (-4.3%)	2.5 (-3.1%)	-	0.44 (4.8%)	-
BII		7467.4 (-0.4)	5366.7 (2.2%)	6.6 (0.6%)	2.7 (3.9%)	0.59 (11.3%)	0.42 (0%)	8141.2 (5.2%)
GBII		7629.2 (1.7%)	5336.4 (1.6%)	6.3 (-2.9%)	2.6 (-0.4%)	1.41 (166%)	0.43 (2.4%)	12794.6 (65.3%)
Empirical	2007-08	10815.2	7500	7.86	2.9	0.50	0.45	10778.1
Exponential		10815.2 (0%)	7496.5 (0%)	21.9 (178%)	4.8 (66.8%)	0.5(0%)	0.5 (11.1%)	11111.1 (3.1%)
Gamma		10815.2 (0%)	8644.8 (15.3%)	9.9 (25.7%)	3.3 (12.8%)	0.79 (58%)	0.41 (-8.9%)	8379.9 (-22.3%)
Weibull		10909.77 (0.9%)	8521.7 (13.6%)	13.2 (68.4%)	3.7 (29.1%)	0.35 (-29.3%)	0.44 (-2.2%)	9174.5 (-14.9%)
Lognormal		11525.7 (6.6%)	7640.0 (1.9%)	8.2 (4.7%)	3.03 (4.9%)	0.48 (-4%)	0.44 (-2.2%)	9102.8 (-15.5)
Fisk		7994.8 (-26.1%)	7547.1 (0.6%)	7.8 (-0.8%)	2.8 (-3.5%)	2.82 (465.9%)	0.47 (4.4%)	26370.4 (144.6%)
Dagum		11734.1 (8.5%)	7416.2 (-1.1%)	7.8 (-0.3%)	2.8 (-2.4%)	-	0.48 (6.7%)	-
Singh-Maddla		11517.0 (6.5%)	7476.4 (-0.3%)	7.8 (-1%)	2.8 (-3.8%)	-	0.49 (8.9%)	-
BII		10980.7 (1.5%)	7496.3 (0%)	8.0 (1.4%)	2.9 (1.4%)	0.78 (56%)	0.46 (2.2%)	13739 (27.5%)
GBII		10868.2 (0.5%)	7517.5 (0.2%)	8.1 (3.4%)	3.0 (3.5%)	0.60 (20%)	0.45 (0%)	11937.6 (10.8%)

Note: Number inside the parenthesis is the percentage difference between empirical index and an index implied by a fitted model. P90/P10 and P75/P25 denote, respectively, the ratio of the 90^{th} percentile to the 10^{th} percentile and the ratio of the 75^{th}percentile to the 25^{th}percentile.

CONCLUSION

The objective of this paper is to model the household income distribution in the province of Punjab Pakistan between the year 2004 to 2008 using parametric income distribution models proposed in the theoretical literature. In particular, we used the four parameters GB2 model and its other eight popular and widely used limiting and nested models on household income dataset using maximum likelihood estimation method. We used log likelihood, BIC, Chi-square, SSE, SAE and likelihood ratio test as a statistical criteria and q-q plot as graphical method which reveal that GB2 is appropriately describing income distribution our both data sets. It is also observed that we can have a relatively big variation in all the four inequality indices Gini, GE(2), and two percentile ratios measurements depending on the choice of the functional form of distribution used, but GB2 model implied indices show performance better than the considered alternative models in close agreement with sample counterparts. The comparative visual and graphical display of empirical and parametric Lorenz curves of all the considered distribution models for both of our data sets is provided. Our findings on inequality imply that surly it has been increased to some extent between the years 2004 to 2008.

Overall, this paper inferred that parametric modeling is a

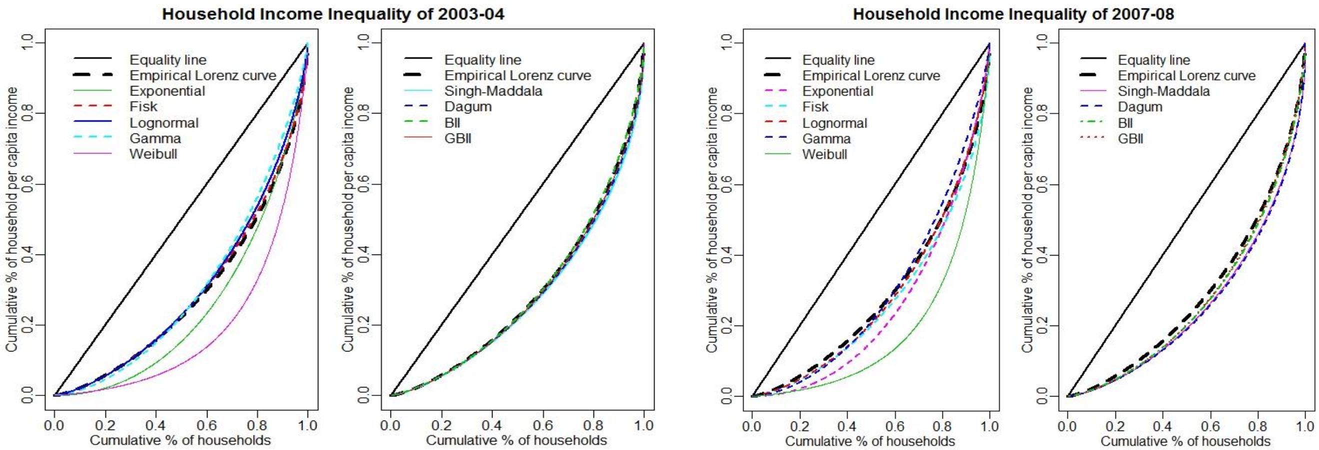

Figure 5. Observed and parametric Lorenz curves for household Income

useful tool to describe the shape, evolution of income distributions and trend in income inequality in the Punjab province, Pakistan.

REFERENCES

1. Chakrabarti, A. S., & Chakrabarti, B. K. (2010). Statistical theories of income and wealth distribution. *Economics: The Open-Access, Open-Assessment E-Journal, 4*, 4.
2. Chotikapanich, D., & Griffiths, W. E. (2008). *Estimating income distributions using a mixture of gamma densities*: Springer.
3. Kakwani, N. (1999). Inequality, welfare and poverty: three interrelated phenomena *Handbook of income inequality measurement* (pp. 599-634): Springer.
4. Pareto, V., & Politique, C. d. E. (1897). Vol. 2. *Rouge, Lausanne*.
5. Boccanfuso, D., Richard, P., & Savard, L. (2013). Parametric and nonparametric income distribution estimators in CGE micro-simulation modeling. *Economic Modelling, 35*, 892-899.
6. Gibrat, R. (1931). *Les inégalités économiques*: Recueil Sirey.
7. Fisk, P. R. (1961). The graduation of income distributions. *Econometrica: journal of the Econometric Society*, 171-185.
8. Salem, A., & Mount, T. (1974). A convenient descriptive model of income distribution: the gamma density. *Econometrica: journal of the Econometric Society*, 1115-1127.
9. Bartels, C., & Van Metelen, H. (1975). Alternative probability density functions of income. *Research memorandum, 29*.
10. Singh, S.K., Maddala, G.S. (1976). "A function for the size distribution of incomes". *Econometrica*, 44, 963–970.
11. Dagum, C. (1977). "A new model for personal income distribution: specification and estimation,"*Economie Appliquée*, 30, 413–437.
12. Taillie, C. (1981). Lorenz ordering within the generalized gamma family of income distributions. *Statistical distributions in scientific work, 6*, 181-192.
13. McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica: journal of the Econometric Society*, 647-663.
14. Bandourian, R., McDonald, J., & Turley, R. S. (2002). A comparison of parametric models of income distribution across countries and over time. Luxembourg Income Study Working Paper No. 305.
15. Brzeziński, M. (2013). Parametric modelling of income distribution in Central and Eastern Europe. *Central European Journal of Economic Modelling and Econometrics, 3*(5), 207-230.
16. Chotikapanich, D., Griffiths, W., Karunarathne, W., & Prasada Rao, D. (2013). Calculating poverty measures from the generalised beta income distribution. *Economic Record, 89*(S1), 48-66.
17. McDonald, J. B., & Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics, 66*(1), 133-152.
18. Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences* (Vol. 470): John Wiley & Sons.
19. Haridas H.N**,** (2007), income modeling using quantile functions, thesis, department of statistics Cochin University of science and technology Cochin-682022.
20. McDonald, J. B., & Ransom, M. (2008). The generalized beta distribution as a model for the distribution of income: estimation of related measures of inequality *Modeling Income Distributions and Lorenz Curves* (pp. 147-166): Springer.
21. Jenkins, S. P. (2009). Distributionally-Sensitive Inequality Indices and The Gb2 Income Distribution. *Review of Income and Wealth, 55*(2), 392-398.
22. Graf, M., Nedyalkova, D., Münnich, R., Seger, J., & Zins, S. (2011). Parametric estimation of income distributions and indicators of poverty and social exclusion. *Research Project Report WP2-D2, 1*.

23. Okamoto, M. (2013). Extension of the κ-generalized distribution: new four-parameter models for the size distribution of income and consumption: LIS working paper 600.
24. Shorrocks, A. F. (1984). "Inequality decomposition by population subgroups," *Econometrica*, 52, 1369–1388.
25. Bresson, F. (2009). On the estimation of growth and inequality elasticities of poverty with grouped data. *Review of Income and Wealth*, *55*(2), 266-302.
26. Thomas W. Yee (2013). VGAM: Vector Generalized Linear and Additive Models. R package version 0.9-3. URL http://CRAN.R-project.org/package=VGAM
27. Graf, M., & Nedyalkova, D. (2012). GB2: Generalized Beta Distribution of the Second Kind: properties, likelihood, and estimation. http://cran.r-project.org/web/packages/GB2.

APPENDIX

Table 2. Distributions and Inequality Measures in Functional form

Distributions	Gini	$GE(2)$	Lorenz curve
Exponential	$G=\frac{1}{2}$	$GE(2)=\frac{1}{2}$	$L(u)=u+(1-u)\ln(1-u)$
Gamma	$G=\frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha+1)\sqrt{\pi}}$	$GE(2)=\frac{1}{2a}$	$(u,L(u))=\left(\frac{\gamma(a,x/\sigma)}{\Gamma(a)},\frac{\gamma(a+1,x/\sigma)}{\Gamma(a+1)},\right)$
Weibull	$G=1-(1/2^{\frac{1}{\alpha}})$	$GE(2)=\frac{-1}{2}+\frac{\Gamma\left(1+\frac{2}{a}\right)}{2*\Gamma^2\left(1+\frac{1}{a}\right)}$	$L(u)=\frac{\gamma\left(\frac{1}{a}+1,-\ln(1-u)\right)}{\Gamma\left(\frac{1}{a}+1\right)}$
Lognormal	$G=2\Phi(\sigma/\sqrt{2})-1$	$GE(2)=\frac{e^{\sigma^2}-1}{2}$	$L(u)=\Phi(\Phi^{-1}(u)-\sigma)$
Fisk	$G=\frac{1}{\alpha}$	$GE(2)=\frac{-1}{2}+\frac{\Gamma\left(1+\frac{2}{a}\right)\Gamma\left(1-\frac{2}{a}\right)}{2*\Gamma^2\left(1+\frac{1}{a}\right)\Gamma^2\left(1-\frac{1}{a}\right)}$	$L(u)=B_1\left(u,1+\frac{1}{a},\frac{a-1}{a}\right)$
Singh-Maddla	$G=1-\frac{\Gamma(q)\Gamma(2q-1/\alpha)}{\Gamma(q-1/\alpha)\Gamma(2q)}$	$GE(2)=-\frac{1}{2}+\frac{\Gamma(q)\Gamma\left(1+\frac{2}{a}\right)\Gamma\left(q-\frac{2}{a}\right)}{2\Gamma^2\left(1+\frac{1}{a}\right)\Gamma^2\left(q-\frac{1}{a}\right)}$	$L(u)=B_1\left(1-(1-u)^{\frac{1}{q}},1+\frac{1}{a},q-\frac{1}{a}\right)$
Dagum	$G=\frac{\Gamma(p)\Gamma(2p+\frac{1}{\alpha})}{\Gamma(2p)\Gamma(p+\frac{1}{\alpha})}-1$	$GE(2)=-\frac{1}{2}+\frac{\Gamma(p)\Gamma\left(p+\frac{2}{a}\right)\Gamma\left(1-\frac{2}{a}\right)}{2\Gamma^2\left(p+\frac{1}{a}\right)\Gamma^2\left(1-\frac{1}{a}\right)}$	$L(u)=B_1\left(u^{\frac{1}{q}},q+\frac{1}{a},1-\frac{1}{a}\right)$
BII	$G=\frac{2B(2p,2q-1)}{PB^2(p,q)}$	$GE(2)=-\frac{1}{2}+\frac{B(p,q)B(p+2,q-2)}{2B^2(p+1,q-1)}$	$L(u)=B_{G2}(B_2^{-1}(u,c,p,q),c,1,p+1,q-1)$