Review of Income and Wealth
Series 55, Number 2, June 2009

DISTRIBUTIONALLY-SENSITIVE INEQUALITY INDICES AND THE GB2 INCOME DISTRIBUTION

by Stephen P. Jenkins*

University of Essex

The Generalized Beta of the Second Kind (GB2) income distribution provides an excellent description of income distributions. However the degree of inequality implied by GB2 parameter estimates is typically summarized using the Gini coefficient only. This paper provides formulae for the Generalized Entropy class of inequality indices for GB2 distributions, thereby providing a full range of top-sensitive and bottom-sensitive measures. The usefulness of having a portfolio of distributionally-sensitive indices is demonstrated using GB2-based estimates of British income inequality in 1994/95 and 2004/05.

1. Introduction

Parametric functional forms have received considerable attention in the literature on earnings and income distribution. They "claim attention, not only for their suitability in modelling some features of many empirical income distributions, but also because of their role as equilibrium distributions in economic processes" (Cowell, 2000, p. 145). Although a large number of functional forms have been proposed, the four-parameter Generalized Beta of the Second Kind (GB2) model is now widely acknowledged to give an excellent description of income distributions, providing fine goodness-of-fit with relative parsimony, while also including many other models as special or limiting cases. See, *inte*r *alia*, Bordley *et al*. (1996), Brachmann *et al*. (1996), Butler and McDonald (1989), McDonald (1984), and McDonald and Xu (1995). Feng *et al*. (2006) address issues of time-inconsistency in topcoded U.S. Current Population Survey earnings data by fitting GB2 distributions that account for topcoding, and derive a consistent time series of Gini coefficients from the estimates. Parker's (1999) model of optimizing firm behavior characterizes an earnings distribution with the GB2 shape.

Despite widespread use of the GB2 distribution, it is remarkable that inequality in the fitted distribution has been summarized in terms of the Gini coefficient alone.[1] Although commonly used, the Gini is but one of many measures of inequality that is available, and it incorporates particular assumptions about the way in which income differences in different parts of the distribution are summa-

Note: Funding support from the U.K. Economic and Social Research Council and the University of Essex for the Research Centre on Micro-Social Change at ISER is gratefully acknowledged. Thanks to Emmanuel Flachaire for suggesting the method to derive $I(0)$ and $I(1)$ and to Fabio Clementi for pointing out some crucial typographical errors in an earlier version. I am also grateful to two anonymous referees, Mark Bryan, Shuaizhang Feng, Lucinda Platt, and Steve Pudney for helpful comments and suggestions.

*Correspondence to: Stephen P. Jenkins, Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ, UK (stephenj@essex.ac.uk).

[1]One exception is Butler and McDonald (1986) who report Pietra ratios as well as Gini coefficients.

rized. (It is relatively sensitive to income differences around the mode.) In other forms of income distribution research, Generalized Entropy (GE) and Atkinson indices are widely used to assess inequality trends and differences—these one-parameter families have the advantage that variations in inequality aversion are straightforwardly incorporated.

This paper provides formulae for GE indices in the GB2 model, and hence also for the important special cases of the three-parameter Singh–Maddala and Dagum models, thereby making a full range of top-sensitive and bottom-sensitive measures available to analysts. The focus is on GE indices because each member of the Atkinson index class has an ordinally equivalent counterpart in the GE class and formulae for Atkinson indices can be derived from their GE counterparts (see below).

The only GE index mentioned in Kleiber and Kotz's (2003) otherwise encyclopaedic survey of the GB2 and related distributions is the Theil index for the Singh–Maddala model. Cowell and Flachaire (2007) provide GE index formulae for the Singh–Maddala model, but using a different parameterization from the standard one that is employed by McDonald (1984) and Kleiber and Kotz (2003). There appear to be no extant GE index formulae for the Dagum distribution, which is surprising given Kleiber's (1996) argument that the Dagum distribution is likely to provide a better fit to income data than the Singh–Maddala distribution.

The usefulness of having a portfolio of distributionally-sensitive indices is demonstrated with an examination of GB2-based estimates of income inequality in Britain in 1994/95 and 2004/05. It is shown that there was a statistically significant increase in inequality according to a top-sensitive GE index, but not according to the Gini coefficient or middle-and bottom-sensitive GE indices.

2. Generalized Entropy Indices

Consider the distribution of a random variable y ("income"), which takes strictly positive values. The generalized entropy (GE) class of inequality measures, $I(\alpha)$, is defined as[2]

$$I(\alpha) = \frac{v_\alpha \mu^{-\alpha} - 1}{\alpha(\alpha - 1)}, \quad \alpha \neq 0, 1 \tag{1}$$

where

$$v_\alpha = \int y^\alpha dF(y), \quad \alpha \neq 0, 1 \tag{2}$$

and $F(y)$ is the cumulative distribution function (cdf) for y. The mean logarithmic deviation (MLD) index is

$$I(0) = \lim_{\alpha \to 0} I(\alpha) = \log \mu - v_0 \tag{3}$$

[2]On the characterization of the GE class of inequality indices, see Bourguignon (1980), Cowell (1980), and Shorrocks (1980, 1984).

where $v_0 = \int \log y dF(y)$ and $\mu \equiv E(y)$ is the mean of y. The Theil index is

$$(4) \qquad I(1) = \lim_{\alpha \to 1} I(\alpha) = \left(\frac{\mu}{v_1}\right) - \log \mu$$

where $v_1 = \int y \log y dF(y)$. $I(2)$ is half the squared coefficient of variation.

Parameter $\alpha \in (-\infty, \infty)$ characterizes the sensitivity of $I(\alpha)$ to income differences in different parts of the income distribution. The more positive that α is, the more sensitive is $I(\alpha)$ to income differences at the top of the distribution; the more negative that α is, the more sensitive is $I(\alpha)$ to income differences at the bottom of the distribution. In empirical work, the range of values for α is typically restricted to [–1, 2] because, otherwise, estimates may be unduly influenced by a small number of very small incomes or very high incomes.

For each member of the Atkinson (1970) class of inequality indices, $A(\varepsilon)$, $\varepsilon > 0$, there is an ordinally equivalent member of the GE class but not vice versa. Specifically, for inequality aversion parameter $\varepsilon = 1 - \alpha$, $\alpha < 1$,

$$(5) \qquad \begin{aligned} A(\varepsilon) &= 1 - [\alpha(\alpha - 1) I(\alpha) + 1]^{\frac{1}{\alpha}}, \quad \alpha < 1, a \neq 0 \\ &= 1 - \exp[-I(0)], \quad \alpha = 0. \end{aligned}$$

Since $A(\varepsilon)$ can be computed from $I(\alpha)$, this paper focuses on the derivation of $I(\alpha)$ in the GB2 distribution case.

3. The GB2 Distribution

The GB2 distribution has probability density function

$$(6) \qquad f(y) = \frac{a y^{ap-1}}{b^{ap} B(p, q) \left[1 + (y/b)^a\right]^{p+q}}, \ y > 0$$

where parameters a, b, p, q, are each positive, $B(u, v) = \Gamma(u)\Gamma(v)/\Gamma(u + v)$ is the Beta function, and $\Gamma(.)$ is the Gamma function (McDonald, 1984). Parameter b is a scale parameter, and a, p, and q are each shape parameters. The k-th moment of the GB2 distribution is

$$(7) \qquad E\left(y^k\right) = \frac{b^k \Gamma\left(p + \frac{k}{a}\right) \Gamma\left(q - \frac{k}{a}\right)}{\Gamma(p)\Gamma(q)}$$

and exists only if $-ap < k < aq$. Tail behavior of the distribution depends on ap (lower tail) and aq (upper tail), with larger values of a reducing the density at both tails, and the relative sizes of p and q affecting skewness (Kleiber and Kotz, 2003).

The Singh–Maddala distribution is the special case of the GB2 distribution when $p = 1$; the Dagum distribution is the special case when $q = 1$. For a discussion of other special cases, see McDonald (1984) and Kleiber and Kotz (2003).

Estimation of the GB2 parameters from unit-record data on incomes is straightforward using maximum likelihood methods. See, for example, Kleiber and Kotz (2003, pp. 193 ff.) for the expression for each log-likelihood contribution, based on the probability density function given in equation (6).[3]

4. GE Inequality Indices and the GB2 Distribution

Expressions for each GE index, $I(\alpha)$, other than for the cases $\alpha = 0, 1$, can be derived by substitution, using the expressions for v_α and μ given by equations (2) and (7). In particular, the bottom-sensitive index $I(-1)$ is given by

$$I(-1) = -\frac{1}{2} + \frac{\Gamma\left(p-\frac{1}{a}\right)\Gamma\left(q+\frac{1}{a}\right)\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)}{2\Gamma^2(p)\Gamma^2(q)}. \tag{8}$$

The top-sensitive index $I(2)$ is given by

$$I(2) = -\frac{1}{2} + \frac{\Gamma(p)\Gamma(q)\Gamma\left(p+\frac{2}{a}\right)\Gamma\left(q-\frac{2}{a}\right)}{2\Gamma^2\left(p+\frac{1}{a}\right)\Gamma^2\left(q-\frac{1}{a}\right)}. \tag{9}$$

Expressions for the more middle-sensitive MLD and Theil indices can be derived noting that the expression for $I(\alpha)$ can be written as $I(\alpha) = g(\alpha)/h(\alpha)$, where $g(\alpha) = v_\alpha \mu^{-\alpha} - 1$, with $v_\alpha = b^\alpha \Gamma\left(p+\frac{\alpha}{a}\right)\Gamma\left(q-\frac{\alpha}{a}\right)\Big/\Gamma(p)\Gamma(q)$ from equation (7), and $h(\alpha) = \alpha(\alpha - 1)$. Hence, using L'Hôpital's rule, $I(0) = -g'(0)$ and $I(1) = g'(1)$, where $g'(\alpha) = (\mu^{-\alpha})' v_\alpha + \mu^{-\alpha}(v_\alpha)'$. From the expressions for $(\mu^{-\alpha})'$ and $(v^\alpha)'$ evaluated at the limits $\alpha \to 0$ and $\alpha \to 1$, it can be

$$I(0) = \gamma\left(p+\frac{1}{a}\right) + \gamma\left(q-\frac{1}{a}\right) - \gamma(p) - \gamma(q) - \frac{\psi(p)}{a} + \frac{\psi(q)}{a} \tag{10}$$

and

$$I(1) = \frac{\psi\left(p+\frac{1}{a}\right)}{a} - \frac{\psi\left(q-\frac{1}{a}\right)}{a} - \gamma\left(p+\frac{1}{a}\right) - \gamma\left(q-\frac{1}{a}\right) + \gamma(p) + \gamma(q) \tag{11}$$

given lngamma function $\gamma(z) = \log\Gamma(z)$ and digamma function $\psi(z) = \Gamma'(z)/\Gamma(z) = \gamma'(z)$.

To derive the expression for $I(\alpha)$ in the special case of the Singh–Maddala model, set $p = 1$ and note that $\Gamma(1) = 1$. For the Dagum model, set $q = 1$ instead.

The expressions for the GE indices show that there is no straightforward relationship between *ceteris paribus* variation in a given parameter and changes in

[3]See McDonald (1984) for the multinomial likelihood expressions appropriate for estimation of GB2 parameters from grouped income data.

index values (except that every index is independent of the scale parameter b). For example, the sign of $\partial I(\alpha)/\partial a$ depends on the values of p and q. Indeed, Kleiber (1999) showed that, for two GB2 distributions A and B, if $a_A \leq a_B$, $a_Ap_A \leq a_Bp_B$, and $a_Aq_A \leq a_Bq_B$, then distribution A Lorenz-dominates distribution B. Necessary conditions for Lorenz dominance are $a_Ap_A \leq a_Bp_B$, and $a_Aq_A \leq a_Bq_B$.

5. Empirical Illustration: Income Inequality in Britain, 1994/95 and 2004/05

The usefulness of having a portfolio of distributionally-sensitive inequality indices for GB2 models is illustrated with analysis of income inequality in Britain. Estimation is based on the unit record data used to calculate the official income statistics, derived from the Family Resources Surveys of fiscal years 1994/95 and 2004/05. "Income" is the distribution among individuals of needs-adjusted post-tax post-transfer household income, with each individual assumed to receive the income of the household to which he or she belongs. Income is net household income before the deduction of housing costs, needs-adjusted using the McClements BHC equivalence scale, and expressed in pounds per week. For further details of the construction of the distributions, see Department for Work and Pensions (2006). Observations with income equal to zero were excluded from the calculations (182 observations in the 1994/95 file and 302 observations in the 2004/05 file).

Estimates of the GB2 parameters for each year are shown in Table 1, together with inequality index estimates implied by them.[4] According to probability plots and quantile plots (not shown), the GB2 distribution fits the data well in each year.

The estimated GB2 shape parameters changed markedly over the decade, with a notable rise in a, combined with a sharp fall in both p and q. Put another way, the distribution was well characterized by a Fisk distribution in 1994/95 (the GB2 case when $p = q = 1$), but could not be described thus a decade later. These changes contrast with the trend in GB2 parameters for 1984–93 reported by Brachmann *et al*. (1996) for household income in Germany, and for 1948–80 for U.S. white family income reported by Butler and McDonald (1989). For both countries, there was a secular decline in a and a rise in p and q.

For Britain, the rise in a combined with a fall in p and in q implies that neither distribution Lorenz-dominates the other one (Kleiber, 1999), so conclusions about whether inequality increased or decreased will, in general, depend on the cardinal inequality index used. As it happens, the GB2 estimates of the Gini coefficient and each of four GE indices increased between 1994/95 and 2004/05, and the increase for the GE indices is greater the more positive that α is. However, of the

[4]A program for fitting a GB2 distribution to unit record data by maximum likelihood methods using the statistical software Stata™ (StataCorp 2003), versions 8.2 and later, is provided by Jenkins (2007). Stata users can install the program directly by typing `ssc install gb2fit`. The maximization algorithm is modified Newton–Raphson (by default), or optionally Berndt–Hall–Hall–Hausman, Davidon–Fletcher–Powell or Broyden–Fletcher–Goldfarb–Shanno. Parameter variances are based on the negative inverse Hessian by default, or optionally OPG. GE and Atkinson inequality indices, and associated standard errors computed using the delta method, can be derived after estimation using the `nlcom` command.

TABLE 1

Estimates of GB2 Parameters and Inequality Indices, Britain, 1994/95 and 2004/05

	1994/95	2004/05
Parameter estimates		
a	2.994	4.257
	(0.057)	(0.179)
b	227.840	341.965
	(4.602)	(4.602)
p	1.063	0.682
	(0.072)	(0.042)
q	1.015	0.635
	(0.058)	(0.037)
Log-likelihood	−196,960	−204,850
N (households)	26,033	25,790
N (individuals)	62,055	59,804
Inequality index estimates		
Gini	0.327	0.329
	(0.003)	(0.003)
I(−1)	0.217	0.220
	(0.005)	(0.005)
I(0)	0.182	0.186
	(0.003)	(0.004)
I(1)	0.198	0.211
	(0.005)	(0.006)
I(2)	0.310	0.397
	(0.017)	(0.030)

Note: Estimated standard errors, shown in parentheses, adjust for within-household clustering.

five indices, it is only for *I*(2)—for which the estimated increase is some 28 percent—that the increase is statistically significant. In this case the test statistic for the relevant *t*-test is 2.5, but it is markedly less than 2 for the other four indices.

The significant rise in top-sensitive index *I*(2) suggests that the principal changes over the decade in the British income distribution occurred at the very top of the distribution. This is confirmed by the GB2 estimates of the Lorenz curves (not shown), which indicate no changes in income shares at the bottom of the income distribution but perceptible increases in income shares at the top. For example, the GB2 estimate of the income share of the richest 5 percent increased from 16.5 to 17.3 percent between 1994/95 and 2004/05, and the income share of the richest 1 percent from 5.6 to 6.3 percent.

If British inequality trends over the decade had been assessed using the Gini coefficient alone, a number of important dimensions of the change would not have been picked up. The ability to calculate a range of indices incorporating different assumptions about aggregation of income differences in different income ranges is a significant extension to the utility of the GB2 model for analysis of income and earnings distributions.

References

Atkinson, Anthony B., "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–63, 1970.

Bordley, Robert F., James B. McDonald, and Anand Mantrala, "Something New, Something Old: Parametric Models for the Size Distribution of Income," *Journal of Income Distribution*, 6, 91–103, 1996.

Bourguignon, François, "Decomposable Inequality Measures," *Econometrica*, 47, 901–20, 1980.

Brachmann, Klaus, Andreas Stich, and Mark Trede, "Evaluating Parametric Income Distribution Models," *Allegemeine Statistiches Archiv*, 80, 285–98, 1996.

Butler, Richard J. and James B. McDonald, "Income Inequality in the U.S.: 1948–80," *Research in Labor Economics*, 8, 85–140, 1986.

———, "Using Incomplete Moments to Measure Inequality," *Journal of Econometrics*, 42, 109–19, 1989.

Cowell, Frank A., "On the Structure of Additive Inequality Measures," *Review of Economic Studies*, 47, 521–531, 1980.

———, "Measurement of Inequality," in Anthony B. Atkinson and François Bourguignon (eds), *Handbook of Income Distribution, Volume 1*, Elsevier, Amsterdam, 87–166, 2000.

Cowell, Frank A. and Emmanuel Flachaire, "Income Distribution and Inequality Measurement: The Problem of Extreme Values," *Journal of Econometrics*, 141, 1044–72, 2007.

Department for Work and Pensions, *Households Below Average Income. An Analysis of the Income Distribution 1994/5–2004/5*, Corporate Document Services, Leeds, 2006 (http://www.dwp.gov.uk/asd/hbai/hbai2005/contents.asp).

Feng, Shuaizhang, Richard V. Burkhauser, and J. S. Butler, "Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution," *Journal of Business and Economic Statistics*, 24, 57–62, 2006.

Jenkins, Stephen P., "`gb2fit`: Stata Module to fit Generalized Beta of the Second Kind Distribution by Maximum Likelihood," Statistical Software Components Archive S456823, 2007 (http://ideas.repec.org/c/boc/bocode/s456823.html).

Kleiber, Christian, "Dagum versus Singh-Maddala Income Distributions," *Economics Letters*, 53, 265–8, 1996.

———, "On the Lorenz Order Within Parametric Families of Income Distributions," *Sankhyā*, B61, 514–17, 1999.

Kleiber, Christian and Samuel Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley, Hoboken, NJ, 2003.

McDonald, James B., "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 52, 647–63, 1984.

McDonald, James B. and Yexiao J. Xu, "A Generalization of the Beta Distribution with Applications," *Journal of Econometrics*, 66, 133–52, 1995 (Erratum: *Journal of Econometrics*, 69, 427–8, 1995).

Parker, Simon C., "The Generalized Beta as a Model for the Distribution of Earnings," *Economics Letters*, 62, 197–200, 1999.

Shorrocks, Anthony F., "The Class of Additively Decomposable Inequality Indices," *Econometrica*, 48, 613–25, 1980.

Shorrocks, Anthony F., "Inequality Decomposition by Population Subgroup," *Econometrica*, 52, 1369–85, 1984.

StataCorp, *Stata Statistical Software: Release 8*, Stata Corporation, College Station, TX, 2003.