



INEQUALITY AND THE GB2 INCOME DISTRIBUTION

Stephen P. Jenkins
(ISER, University of Essex)

ISER Working Paper
2007-12

Acknowledgement

ISER's core-funding from the Economic and Social Research Council and the University of Essex is gratefully acknowledged. Thanks to Emmanuel Flachaire for suggesting the method to derive $I(0)$ and $I(1)$, and to Mark Bryan, Shuaizhang Feng, Lucinda Platt, and Steve Pudney for comments.

Fabio Clementi (Polytechnic University of Marche) pointed out typos in equations (13) and (14) in the original version of the working paper. These have been now been corrected (May 2008).

Readers wishing to cite this document are asked to use the following form of words:

Jenkins, Stephen P. (May 2007). 'Inequality and the GB2 income distribution', ISER Working Paper 2007-12. Colchester: University of Essex. (Revised May 2008)

The on-line version of this working paper can be found at <http://www.iser.essex.ac.uk/pubs/workpaps/>

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates

- MISOC (the ESRC Research Centre on Micro-social Change), an international centre for research into the lifecourse, and
- ULSC (the ESRC UK Longitudinal Studies Centre), a national resource centre to promote longitudinal surveys and longitudinal research.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Institute for Social and Economic Research, University of Essex, Wivenhoe Park,
Colchester. Essex CO4 3SQ UK
Telephone: +44 (0) 1206 872957 Fax: +44 (0) 1206 873151 E-mail: iser@essex.ac.uk
Website: <http://www.iser.essex.ac.uk>

© May 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form, or by any means, mechanical, photocopying, recording or otherwise, without the prior permission of the Communications Manager, Institute for Social and Economic Research.

ABSTRACT

The generalized entropy class of inequality indices is derived for Generalized Beta of the Second Kind (GB2) income distributions, thereby providing a full range of top-sensitive and bottom-sensitive measures. An examination of British income inequality in 1994/95 and 2004/05 illustrates the analysis.

Key words: inequality, generalized entropy indices, generalized Beta of the second kind distribution, GB2 distribution, Singh-Maddala distribution, Dagum distribution

JEL classifications: C16, C46, D31.

NON-TECHNICAL SUMMARY

An important strand of the literature that aims to summarize earnings and income distributions chooses to summarise the shapes of these distributions in mathematical terms. The key features of the distribution are encapsulated in a formula which contains a small number of parameters that can be estimated from survey data.

Readers may be familiar with the symmetric bell-shaped curve that describes the Normal distribution, the formula for which has two parameters (the mean and the variance). However, this shape is inappropriate for describing most income distributions because they tend to be skewed, with a peak in the lower-middle income range and to have a long right-hand tail. To capture these features, a large number of functional forms other than the Normal distribution have been proposed. Of these, 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.

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. Although this index is commonly used in distributional analysis of all kinds, it is but one of many measures available, and it incorporates particular assumptions about how income differences at different points along the income range are summarized. (The Gini is relatively sensitive to income differences around the mode.) In other forms of distribution analysis, researchers commonly use generalized entropy (GE) and Atkinson indices 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 generalized entropy 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 paper's derivations of expressions for GE inequality indices in the context of the GB2 model are illustrated with an examination of the change in income inequality in Britain between 1994/95 and 2004/05, using exactly the same data as used in the British official income statistics (the so-called 'HBAI' data). The estimates reveal that income inequality rose between 1994/95 and 2004/05 according to the Gini coefficient and four GE inequality measures. However the increase was statistically significant only for the top-sensitive inequality index $I(2)$, which 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, which indicate imperceptible changes in income shares at the bottom of the income distribution but increases in income shares at the top. For example, the GB2 estimate of the income share of the richest five per cent increased from 16.5 per cent to 17.3 per cent between 1994/95 and 2004/05, and the income share of the richest one per cent from 5.6 per cent to 6.3 per cent.

If British inequality trends over the decade had been assessed using the GB2-estimated 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.

1 Introduction

Parametric functional forms have received considerable attention in the literature on earnings and income distribution. 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, *inter alia*, Bordley et al. (1999), Brachmann et al. (1999), Butler and McDonald (1989), McDonald (1984), and McDonald and Xu (1995). Feng et al. (2006) address issues of time-inconsistency in top-coded US Current Population Survey earnings data by fitting GB2 distributions that account for top-coding, and derive a consistent time series of Gini coefficients from the estimates. Parker's (1999) model of optimising firm behaviour predicts that the earnings distribution has the GB2 shape.

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

Since each member of the Atkinson index class has an ordinally equivalent counterpart (see below), the principal focus of the paper is on GE indices, and formulae for Atkinson indices are derived from them. The only GE index mentioned in Kleiber and Kotz's (2003) 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 sur-

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

prising given Kleiber's (1996) argument that the Dagum distribution is likely to provide a better fit to income data than the Singh-Maddala distribution. This paper's derivations for the GB2 model are illustrated with an examination 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 as²

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

where

$$\nu_\alpha = E(y^\alpha) = \int y^\alpha dF(y), \quad \alpha \neq 0, 1. \quad (2)$$

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

$$I(0) = \lim_{\alpha \rightarrow 0} I(\alpha) = \log \mu - \nu_0 \quad (3)$$

where $\nu_0 = \int \log y dF(y)$ and $\mu \equiv E(y)$ is the mean of y . The Theil index is

$$I(1) = \lim_{\alpha \rightarrow 1} I(\alpha) = \left(\frac{\mu}{\nu_1} \right) - \log \mu \quad (4)$$

where $\nu_1 = \int y \log y dF(y)$. $I(2)$ 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.

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

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

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

Since $A(\epsilon)$ 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

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

where parameters a, b, p, q , are each positive, $B(u, v) = \Gamma(u)\Gamma(v)/\Gamma(u + v)$ is the Beta function, and $\Gamma(\cdot)$ 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

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

and exists only if $-ap < k < aq$.

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 e.g. Kleiber and Kotz (2003, pp. 193 ff.) for the expression for each log-likelihood contribution, based on the probability density (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 ν_α and μ given by (2)

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

and (7). In particular, the bottom-sensitive index $I(-1)$ is given by

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

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

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

Expressions for the more middle-sensitive MLD and Theil indices can be derived noting that the expression for $I(\alpha)$ can be written as

$$I(\alpha) = \frac{g(\alpha)}{h(\alpha)} \quad (10)$$

where $g(\alpha) = \nu_\alpha \mu^{-\alpha} - 1$, with $\nu_\alpha = b^\alpha \Gamma(p + \frac{\alpha}{a})\Gamma(q - \frac{\alpha}{a})/\Gamma(p)\Gamma(q)$ from (7), and $h(\alpha) = \alpha(\alpha - 1)$. Hence, using L'Hôpital's rule, $I(0) = -g'(0)$ and $I(1) = g'(1)$, where $g'(\alpha) = (\mu^{-\alpha})'\nu_\alpha + \mu^{-\alpha}(\nu_\alpha)'$. We therefore require expressions for $(\mu^{-\alpha})'$ and $(\nu_\alpha)'$ evaluated at the limits $\alpha \rightarrow 0$, and $\alpha \rightarrow 1$. It can be shown that

$$(\mu^{-\alpha})' = -\mu^{-\alpha} \log \mu \quad (11)$$

and

$$(\nu_\alpha)' = \nu_\alpha \left[\frac{\psi(p + \frac{\alpha}{a})}{a} - \frac{\psi(q - \frac{\alpha}{a})}{a} + \log b \right] \quad (12)$$

given digamma function $\psi(z) = \Gamma'(z)/\Gamma(z)$. Hence, applying the appropriate limits,

$$I(0) = \gamma(p + \frac{1}{a}) + \gamma(q - \frac{1}{a}) - \gamma(p) - \gamma(q) - \frac{\psi(p)}{a} + \frac{\psi(q)}{a} \quad (13)$$

and

$$I(1) = \frac{\psi(p + \frac{1}{a})}{a} - \frac{\psi(q - \frac{1}{a})}{a} - \gamma(p + \frac{1}{a}) - \gamma(q - \frac{1}{a}) + \gamma(p) + \gamma(q) \quad (14)$$

given lngamma function $\gamma(z) = \log \Gamma(z)$.

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.

5 Empirical illustration: income inequality in Britain, 1994/95 and 2004/05

The derivations are illustrated with analysis of income inequality in Britain in fiscal years 1994/95 and 2004/05. Estimation is based on the unit record data used to calculate the official income statistics, derived from the Family Resources Surveys of 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 s/he belongs. In UK jargon, 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 1984–1993 reported by Brachmann et al. (1996) for household income in Germany, and for 1948–1980 for US white family income reported by Butler and McDonald (1989). In both cases, there was a secular decline in a and a rise in p and q .

<Table 1 near here>

⁴A program for fitting a GB2 distribution to unit record data by maximum likelihood methods using the statistical software StataTM (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.

The rise in a combined with a fall in p and q implies that neither distribution Lorenz-dominates the other one (Kleiber 1999), so conclusions about whether inequality increased or decreased depend on the 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 five indices, it is only for $I(2)$ – for which the estimated increase is some 28 per cent – 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 imperceptible changes in income shares at the bottom of the income distribution but increases in income shares at the top. For example, the GB2 estimate of the income share of the richest five per cent increased from 16.5 per cent to 17.3 per cent between 1994/95 and 2004/05, and the income share of the richest one per cent from 5.6 per cent to 6.3 per cent.

If British inequality trends over the decade had been assessed using the GB2-estimated 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.

6 Acknowledgements

ISER's core-funding from the Economic and Social Research Council and the University of Essex is gratefully acknowledged. Thanks to Emmanuel Flachaire for suggesting the method to derive $I(0)$ and $I(1)$, and to Mark Bryan, Shuaizhang Feng, Lucinda Platt, and Steve Pudney for comments.

7 References

Atkinson, A.B., 1970, On the measurement of inequality, *Journal of Economic Theory* 2, 244–263.

Bordley, R.F., J.B. McDonald and A. Mantrala, 1996, Something new, something old: parametric models for the size distribution of income, *Journal*

of Income Distribution 6, 91–103.

Bourguignon, F. 1989, Decomposable inequality measures, *Econometrica* 47, 901–920.

Brachmann, K., A. Stich and M. Trede, 1996, Evaluating parametric income distribution models, *Allgemeine Statistisches Archiv* 80, 285–298.

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

Butler, J.B. and J.B. McDonald, 1989, Using incomplete moments to measure inequality, *Journal of Econometrics* 42, 109–119.

Cowell, F.A., 1980, On the structure of additive inequality measures, *Review of Economic Studies* 47, 521–531.

Cowell, F.A. and E. Flachaire, 2007, Income distribution and inequality measurement: the problem of extreme values, *Journal of Econometrics*, forthcoming.

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

Feng, S., R.V. Burkhauser and J.S. Butler, 2006, 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.

Jenkins, S.P., 2007, `gb2fit`: Stata module to fit Generalized Beta of the Second Kind distribution by maximum likelihood, *Statistical Software Components Archive S456823*. <http://ideas.repec.org/c/boc/bocode/s456823.html>

Kleiber, C., 1996, Dagum versus Singh-Maddala income distributions, *Economics Letters* 53, 265–268.

Kleiber, C., 1999, On the Lorenz order within parametric families of income distributions, *Sankhyā B* 61, 514–517.

Kleiber, C. and S. Kotz, 2003, *Statistical Size Distributions in Economics and Actuarial Sciences*. (John Wiley, Hoboken, NJ).

McDonald, J.B., 1984, Some generalized functions for the size distribution of income, *Econometrica* 52, 647–663.

McDonald, J.B. and Y.J. Xu, 1995, A generalization of the beta distribution with applications, *Journal of Econometrics* 66, 133–152. Erratum: *Journal of Econometrics* 69, 427–428.

Parker, S.C., 1999, The generalized beta as a model for the distribution of earnings, *Economics Letters* 62, 197–200.

Shorrocks, A.F., 1980, The class of additively decomposable inequality indices, *Econometrica* 48, 613–625.

Shorrocks, A.F., 1984, Inequality decomposition by population subgroup, *Econometrica* 52, 1369–1385.

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

Estimated standard errors shown in parentheses.

Table 1: Estimates of GB2 parameters and inequality indices, Britain, 1994/95 and 2004/05.

StataCorp, 2003, Stata Statistical Software: Release 8.0. (Stata Corporation, College Station TX).