Queen's Economics Department Working Paper No. 913

Estimation and Inference for Normative Inequality Indices

Ian Rongve Charles Beach

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

11-1994

Discussion Paper #913

Estimation and Inference for Normative Inequality Indices

by

Charles M. Beach
Queen's University

Ian Rongve
University of Regina

November 1994

Estimation and Inference for Normative Inequality Indices

Ian Rongve
Department of Economics
University of Regina
Regina, SK.
Canada

and

Charles Beach
Department of Economics
Queen's University
Kingston, ON.
Canada

November 1994

Abstract

This paper addresses the sampling distribution of two general classes of normative income inequality indices. Specifically, for approximations to the Atkinson–Kolm–Sen index of relative inequality and the Kolm–Blackorby-Donaldson index of absolute inequality, the paper provides consistent estimators based on a random sample of income microdata and establishes their asymptotic normality. Asymptotic variance–covariance expressions are obtained that provide for distribution–free statistical inference on these measures in a straightforward fashion. The paper thus extends the principles of statistical inference to current classes of normative inequality measures and to Atkinson's equally distributed income measure, using a general approximation approach. An example using Canadian family income microdata illustrates the technique.

1 Introduction

Applied welfare economists investigating the income inequality present in a population typically select a sample of individuals and compute an estimate of the population inequality index from this sample. Traditionally the point estimates of the index were all that was considered with no attention paid to the sampling accuracy of the estimates. Recently, however, much attention has been focused on the statistical properties of the estimates of these indices. In this paper, we examine the estimation and the asymptotic distribution of two popular classes of inequality indices using a general approximation approach.

The families of indices we consider are the Atkinson-Kolm-Sen index of inequality and the Kolm-Blackorby-Donaldson index of inequality. The analysis in this paper builds on work by Beach and Davidson (1983), Beach and Richmond (1985), Beach, Formby, Chow, and Slotsve (1994). Closely related is work by Barrett (1993), Cowell (1989), Thistle (1990) and Gastwirth (1974) who all derive the sampling distribution of a specific index, whereas the current paper provides an approach applicable to a wide range of inequality measures.

The analysis presented in this paper has its perspective in the literature on ethical or normative inequality indices such as Blackorby and Donaldson (1978) or Chakravarty (1990). This is done primarily for ease of exposition, but the results do not depend on this perspective. The indices considered can also be interpreted as strictly statistical measures. The approximation approach used in this paper has several advantages. First, the results can be applied directly to a large class of inequality indices instead of the case–

by–case nature of previous work in this area. Second, this paper clearly brings out some of the statistical foundations of social welfare measurement, complementing work on the ethical foundations that has been done by welfare theorists. Third, the formulas for the asymptotic standard errors turn out to be 'distribution free' in that they do not depend on the underlying income distribution function having a specific form.

2 Statistical Preliminaries

The material in the next sections builds on work done by Beach and Davidson (1983) and Beach, Formby, Chow, and Slotsve (1994). In this section we present their results as an introduction to our own.

Suppose we have a distribution of income y, $y \in [y_l, y_u]$ from the continuous distribution function $F : [y_l, y_u] \rightarrow [0, 1]$ which has the density function f. For convenience assume that $y_l \geq 0$ so that all incomes are non-negative. Assume that F is strictly increasing in y and invertible. Corresponding to each ordinate p of the distribution F is a value of income, y_p, for which $F(y_p) = p$. Since the distribution function F is invertible, there is a function $G = F^{-1}$ such that $G(p) = y_p$.

Now we divide the possible realizations of y into $S + 1$ equally probable[1] regions with bounds $\{p_i \mid i = 1, ..., S\}$ where $0 < p_i < p_{i+1} < 1$. Also define $\bar{p}_i = (p_i + p_{i-1})/2$. Corresponding to each p_i is an element of the possible realizations of y denoted by the quantile level $\xi_i = G(p_i)$. Define the ith quantile mean to be the conditional expectation of y, given that the

[1] That the regions are equally probable is not required for the distributional results.

realization is in the ith quantile interval. It is

$$\mu_i = \int_{p_{i-1}}^{p_i} \frac{G(p)dp}{p_i - p_{i-1}} \tag{1}$$

with the convention that $p_0 = 0$ and $p_{S+1} = 1$. This is simply the expected income of an individual who is in the ith quantile income interval.

To get a sample estimate of the values of ξ_i, we take a random sample of N observations from the income distribution. Order these from smallest to largest so that given the ordered observations $y_{(i)}$, $i = 1,..N$, we have $y_{(i)} \leq y_{(i+1)}$. Define $r_i = [Np_i]$ to be the greatest integer less than or equal to Np_i. The statistic $\hat{y}_i = y_{(r_i)}$, where $y_{(r_i)}$, the r_ith observation in the ordered sample, is an estimator of ξ_i. Thus an estimator of μ_k is

$$\hat{\mu}_k = \sum_{j=r_{k-1}+1}^{r_k} \frac{y_{(j)}}{r_k - r_{k-1}} \tag{2}$$

Now define the population *cumulative* means to be

$$\gamma_i = \int_0^{p_i} \frac{G(p_i)}{p_i} dp \tag{3}$$

This is the mean of the incomes conditional on being in the lower p_i of the distribution. The sample cumulative means are given by

$$\hat{\gamma}_i = \sum_{j=1}^{r_i} \frac{y_{(j)}}{r_i} \tag{4}$$

which is the average of the lowest r_i observations of income. Now define the vector $M = [\,\mu_1 \quad \mu_2 \quad \ldots \quad \mu_{S+1}\,]$ to be the vector of quantile means and the vector $\hat{M} = [\,\hat{\mu}_1 \quad \hat{\mu}_2 \quad \ldots \quad \hat{\mu}_{S+1}\,]$ to be the sample estimates of M. The vector $\hat{M}$ can be obtained from the vector of cumulative means $\hat{\Gamma}$ by the following linear transformation $\hat{M} = Q\hat{\Gamma}$, where

$$Q = \begin{bmatrix} \frac{p_1}{(p_1-p_0)} & 0 & \cdots & & 0 \\ \frac{-p_1}{(p_2-p_1)} & \frac{p_2}{(p_2-p_1)} & \cdots & & 0 \\ 0 & & & & \\ & & & & \\ 0 & \cdots & 0 & \frac{-P_S}{(P_{S+1}-P_S)} & \frac{P_{S+1}}{(P_{S+1}-P_S)} \end{bmatrix} \tag{5}$$

Beach *et al* (1994) derive the following theorem that gives the asymptotic distribution of the vector of sample quantile means.

Theorem 1 *Under the conditions that the population has a finite mean and variance, the C.D.F. F is strictly monotonic and twice differentiable, the vector* $N^{1/2}[\hat{\mathrm{M}} - \mathrm{M}]$ *is asymptotically normal with mean 0 and covariance matrix* V *given by* $V = R\Omega R'$ *where* R *is defined as*

$$R = \begin{bmatrix} \frac{1}{(p_1-p_0)} & 0 & \cdots & & 0 \\ \frac{-1}{(p_2-p_1)} & \frac{1}{(p_2-p_1)} & \cdots & & 0 \\ 0 & & & & \\ & & & & \\ 0 & \cdots & 0 & \frac{-1}{(P_{S+1}-P_S)} & \frac{1}{(P_{S+1}-P_S)} \end{bmatrix} \tag{6}$$

and Ω *is the symmetric matrix defined by*

$$\omega_{ij} = p_i[{\lambda_i}^2 + (1-p_i)(\xi_{p_i}-\gamma_i)(\xi_{p_j}-\gamma_j) + (\xi_{p_i}-\gamma_i)(\gamma_j-\gamma_i)], \qquad \text{for} \quad i \leq j \tag{7}$$

Here ${\lambda_i}^2$ *is the variance of* y*, conditional on* $y \leq G(p_i)$*.*

Doing the multiplication, we get the i,jth element of V,

$$v_{ij} = \frac{\omega_{ji} - \omega_{j-1i} - \omega_{i-1j} + \omega_{j-1i-1}}{(p_j - p_{j-1})(p_i - p_{i-1})} \tag{8}$$

for $i \leq j$ with the convention that $\omega_{i0} = 0$, $\omega_{0j} = 0$, and $p_0 = 0$.

3 Estimation of Normative Inequality Measures

In this section we describe the ethical approach to measuring inequality, the social preferences over income distributions on which the analysis is based, and the statistical approach used to estimate the inequality indices. The key novelty is the approximation approach introduced in equation (17).

Suppose that society has preferences over alternative income distributions represented by the equally distributed equivalent income (EDE) functional $e = \bar{\mathcal{E}}(F)$. The EDE gives a measure of the amount of per capita income that could provide society with the same level of welfare that it currently enjoys, provided that the income was distributed equally. Now define the Atkinson-Kolm-Sen (AKS) index of relative inequality to be

$$I = 1 - \frac{e}{\mu} \tag{9}$$

and the Kolm-Blackorby-Donaldson (KBD) index of absolute inequality to be

$$A = \mu - e \tag{10}$$

where $\mu = \int_{y_l}^{y_u} y dF(y)$ is the expected value of income from the distribution F.

The AKS relative index defines inequality to be the percentage of potential welfare in the economy that is lost as a result of inequality in the given income distribution. The KBD index gives a money measure of the per capita welfare lost as a result of inequality.

To be more specific, consider the form of the EDE function given by the following,

$$e = \mathcal{E}\left[\int_{y_l}^{y_u} \bar{\psi}_y(y) f(y) dy\right]. \tag{11}$$

That is, the EDE is some transform of the expectation of a given function of income. The term within the square brackets in equation (11) is of interest because this is the way in which society's preferences regarding inequality are introduced. Without the function $\bar{\psi}_y$, the inside term will just be expected income and the EDE income will just be some transform of mean income. In general the function $\bar{\psi}_y$ will be a concave function of y and the function $\mathcal{E}$ will be used to rescale the units of the EDE into dollars. Note that in this formulation the function of income is allowed to vary with income. This allows us to include, as a special case, EDE's such as the S-Gini which depend on the rank order of income. However, we are not restricted to rank orders in this formulation as $\bar{\psi}_y$ may also be constant over values of y.

The formula in equation (11) has a number of continuous EDE functions as special cases. For example, consider

$$\bar{\psi}_y = \delta(1 - F(y))^{\delta - 1} y \tag{12}$$

and

$$\mathcal{E}(x) = x. \tag{13}$$

The resulting EDE function is the Donaldson and Weymark (1983) extension of the S-Gini EDE to the income continuum. The AKS index constructed from the S-Gini has as a special case the well known Gini index of inequality. The parameter δ in equation (12) is an inequality aversion parameter which can vary between one and infinity. At $\delta = 1$ the S-Gini has no inequality aversion, and thus the above inequality indices will be zero. As δ approaches infinity the EDE approaches the income of the least wealthy individual and the inequality index then depends only on the lowest income and the mean income. The standard Gini coefficient is yielded when $\delta = 2$.

The S-Gini EDE function is not the only one with varying inequality aversion. As another example, consider

$$\bar{\psi}_y = y^r \tag{14}$$

and

$$\mathcal{E}(x) = x^{1/r}. \tag{15}$$

We now have a mean of order r EDE function from Blackorby and Donaldson (1978). In this case, the parameter r is the inequality aversion parameter and can vary between one and negative infinity. Again at $r = 1$ the EDE displays no inequality aversion, while as r approaches negative infinity, the EDE again approaches the income of the least wealthy individual. The limit as r approaches zero is the Cobb–Douglas EDE function.

Changing the variables of integration in (11) by use of the inverse function G and recognizing that $G'(p) = 1/F'(G(p))$, yields the following equivalent form of the EDE function

$$e = \mathcal{E}\Big[\int_0^1 \psi_p(G(p))dp\Big] \tag{16}$$

where $\psi_p = \bar{\psi}_{G(p)}$.

While theoretically appealing, the formulation in (16) is not estimable without knowing the exact form of the distribution function F. The objective in the rest of this section is to develop a suitable approximation to (16) that can be consistently estimated from a microdata sample of income without knowledge of the actual distribution function F; *ie.*, a 'distribution-free' method in the sense of Beach and Davidson (1983).

Now consider the following approximation to (16):

$$\tilde{e} = \mathcal{E}\Big[\sum_{k=1}^{S+1} \frac{\psi_{\bar{p}_k}(\mu_k)}{S+1}\Big]. \tag{17}$$

This approximation is constructed by averaging the values of the function $\psi_{\bar{p}_k}$ evaluated at the k'th quantile mean and substituting this average in for the term in square brackets in equation (11). By doing this we lose a bit of the inequality aversion in equation (11), but the benefits in being able to determine the sampling distribution of the EDE estimate (and inequality measures based on it) outweigh the costs[2]. The following theorem shows that as the number of quantiles increases, the approximation in equation (17) gets better in the sense of coming closer to equation (11).

Theorem 2 *Consider the approximation defined in (17) and let the number of quantiles considered approach ∞. Then $\lim_{S\to\infty} \tilde{e} = e$.*

Proof: Since each of the quantiles is equally probable we have $p_i - p_{i-1} = 1/(S+1)$ and $\bar{p}_i = p_i - 1/(2(S+1))$. Then consider

$$\lim_{S\to\infty} \mu_i = \lim_{S\to\infty} \int_{p_i - \frac{1}{S+1}}^{p_i} \frac{(G(p))dp}{1/(S+1)} \tag{18}$$

which by L'Hôpital's rule is

$$\lim_{S\to\infty} \frac{G(p_i - 1/(S+1))(S+1)^2}{(S+1)^2} = G(p_i) \tag{19}$$

and so we have

$$\lim_{S\to\infty} \psi_i(\mu_i) = \lim_{S\to\infty} \psi_{p_i - \frac{1}{2(S+1)}}(G(p_i)) = \psi_{p_i}(G(p_i)) \tag{20}$$

[2]This is not as big a restriction as it may seem as the number of quantiles considered can be as large as desired, given typically large microdata sets. This allows the researcher, given the availability of large numbers of data, to get arbitrarily close to estimating the true function in equation (11).

and thus

$$\lim_{S\to\infty} \tilde{e} = \mathcal{E}\Big[\lim_{S\to\infty}\Big[\sum_{k=1}^{S+1} \frac{\psi_{p_i}(G(p_i))}{S+1}\Big]\Big] \tag{21}$$

Thus

$$\lim_{S\to\infty} \tilde{e} = \mathcal{E}\Big[\int_0^1 \psi_p(G(p))dp\Big] = e \tag{22}$$

which establishes the proposition. **QED**

Equation (17) now provides the basis for the distribution–free method that we use to estimate the EDE. Instead of estimating (16) directly we will estimate the approximation given in (17). Note that this approximation is identical to the lower bound approximation for Gastwirth's (1975) estimation with grouped data[3]. Gastwirth's upper bound approximation, while useful in evaluating the size of grouping error, when applied to microdata–based quantiles is not distribution–free in the sense of this paper.

The corresponding estimator of e that will be used is

$$\hat{e} = \mathcal{E}\Big[\sum_{k=1}^{S+1} \frac{\psi_{\bar{p}_k}(\hat{\mu}_k)}{S+1}\Big]. \tag{23}$$

Now all that remains is an estimate of the expected income from the distribution F. This is given by

$$\hat{\mu} = \sum_{i=1}^{N} \frac{y_{(i)}}{N} = \frac{\sum_{k=1}^{S+1} \hat{\mu}_k}{S+1} = \sum_{i=1}^{N} \frac{y_i}{N}. \tag{24}$$

Given $\hat{e}$ and $\hat{\mu}$, the corresponding sample estimates of the AKS and KBD inequality indices are respectively

$$\hat{I}(y) = 1 - \frac{\hat{e}}{\hat{\mu}} \tag{25}$$

[3]See also Cowell (1977), chapter 5.

and

$$\hat{A}(y) = \hat{\mu} - \hat{e}. \tag{26}$$

4 Inference with Inequality Indices

In this section, we show that the proposed estimators for the AKS index and the KBD index are consistent estimates when based on the approximation equation (17). We also derive the asymptotic distribution of the proposed estimators.

Given the (asymptotic) distribution of the vector $\hat{\text{M}}$, it is easy to derive the (asymptotic) distribution of the inequality indices $\hat{I}$ and $\hat{A}$. This result is obtained in two steps; first we derive the joint distribution of the estimates of the EDE and the mean income $(\hat{e}, \hat{\mu})$. This result is then used to derive the distribution of the indices themselves ($\hat{I}$ and $\hat{A}$).

First define the vector of partial derivatives of $\mathcal{E}$ with respect to the quantile means as

$$d = \begin{bmatrix} \frac{\partial \hat{e}}{\partial \hat{\mu}_1} & \frac{\partial \hat{e}}{\partial \hat{\mu}_2} & \cdots & \frac{\partial \hat{e}}{\partial \hat{\mu}_{S+1}} \end{bmatrix} \tag{27}$$

and let s be the $S+1$ vector with $1/(S+1)$ in each element.

Theorem 3 *The vector*

$$N^{1/2} \begin{bmatrix} \hat{e} - \tilde{e} \\ \hat{\mu} - \mu \end{bmatrix} \tag{28}$$

is distributed asymptotically normal with mean 0 and covariance matrix given by

$$\Phi = \begin{bmatrix} d \\ s \end{bmatrix} V \begin{bmatrix} d \\ s \end{bmatrix}' \tag{29}$$

where $[\,d \quad s\,]'$ *is a 2 by* $S+1$ *matrix.*

Proof: See Rao (1965), Page 322.**QED**

Theorem 3 makes it easy to derive the asymptotic distribution of the indices, but it is also of some interest in itself. e is an empirical representation of the average economic well–being in a population. When doing social cost–benefit analysis of a policy that has both aggregate and distributional impacts, such as unemployment insurance, one of the usual methods is to do a social net present value of the social policy, comparing the change in the EDE (the benefit) with the per-capita cost of the policy over some time horizon[4]. Theorem 3 provides a way of determining how much confidence we should have in the results of this net present value calculation. For example, confidence intervals can be constructed for the net present value using Theorem 3 and information about the variance of the costs of the policy.

The distribution of the inequality indices can now be calculated as

Theorem 4 *Defining $\tilde{I}(y) = 1 - \tilde{e}/\mu$, the estimated AKS inequality index has an asymptotic distribution*

$$N^{1/2}[\hat{I}(y) - \tilde{I}(y)] \xrightarrow{\mathrm{d}} N(0, [\,1/\mu, \quad -e/\mu^2\,]\,\Phi\,[\,1/\mu, \quad -e/\mu^2\,]'); \tag{30}$$

and defining $\tilde{A}(y) = \mu - \tilde{e}$, the KBD absolute inequality index has a limiting distribution

$$N^{1/2}[\hat{A}(y) - \tilde{A}(y)] \xrightarrow{\mathrm{d}} N(0, [\,-1, \quad 1\,]\,\Phi\,[\,-1, \quad 1\,]') \tag{31}$$

Proof: The proof from Rao (1965), page 321, applies. **QED.**

[4]See, for example, Layard and Zabalza (1979).

Defining

$$g_I = [1/\mu, \quad -e/\mu^2] \begin{bmatrix} d \\ s \end{bmatrix} \tag{32}$$

the variance of the AKS index can be calculated directly from the covariance matrix of the quantile means by the formula $g_I V g_I'$. The formula in the above theorem has as two special cases the results for the Gini index in Beach and Davidson (1983) as well as the result for the S-Gini from Barrett (1993).

5 Empirical Illustration

In this section we provide an empirical example of how the estimation described in the previous sections can be applied to practical distribution problems. The specific inequality indices that are used are the S-Gini index of Donaldson and Weymark (1983) and the mean of order r index.

The S-Gini EDE function is defined by equations (11)–(13). The resulting equation for the approximation $\tilde{e}$ is

$$\tilde{e} = \sum_{k=1}^{S+1} \frac{\delta(1-\bar{p}_k)^{\delta-1}\mu_k}{S+1}. \tag{33}$$

In the case of deciles we have $S+1 = 10$ and

$$\tilde{e} = \frac{\delta .95^{\delta-1}\mu_1 + \delta .85^{\delta-1}\mu_2 + \delta .75^{\delta-1}\mu_3 + ... + \delta .05^{\delta-1}\mu_{10}}{10}. \tag{34}$$

To get the sample estimate $\hat{e}$, the quantile means $\hat{\mu}_k$ are substituted into the expression for $\tilde{e}$. The estimate of the AKS inequality index is obtained from equation (25). The variance of $\hat{I}$ is obtained with the $\mathbf{g_I}$ vector defined as

$$\mathbf{g_I} = \left[\frac{\delta .95^{\delta-1}}{10\mu} - \frac{e}{10\mu^2}, \quad \frac{\delta .85^{\delta-1}}{10\mu} - \frac{e}{10\mu^2}, \quad ..., \quad \frac{\delta .05^{\delta-1}}{10\mu} - \frac{e}{10\mu^2} \right] \tag{35}$$

Now the mean of order r EDE function is defined by equations (11), (14), and (15). The resulting equation for the approximation $\tilde{e}$ is

$$\tilde{e} = \left[\sum_{k=1}^{S+1} \frac{\mu_k{}^r}{S+1}\right]^{1/r} \tag{36}$$

In the case of deciles this reduces to

$$\tilde{e} = \left[\frac{\mu_1{}^r + \mu_2{}^r + \ldots + \mu_{10}{}^r}{10}\right]^{1/r} \tag{37}$$

Again use the sample estimate $\hat{e}$ in equation (25) to get the estimate of the inequality index.

Now to obtain the variance of $\hat{I}$ use the $\mathbf{g_I}$ vector defined by

$$\mathbf{g_I} = \left[\frac{r\mu_1{}^{r-1}e^{\frac{1-r}{r}}}{10\mu} - \frac{e}{10\mu^2},\quad \frac{r\mu_2{}^{r-1}e^{\frac{1-r}{r}}}{10\mu} - \frac{e}{10\mu^2},\quad \ldots,\quad \frac{r\mu_{10}{}^{r-1}e^{\frac{1-r}{r}}}{10\mu} - \frac{e}{10\mu^2}\right] \tag{38}$$

The indices were calculated for samples of economic families from the 1992 Family Expenditure Survey compiled by Statistics Canada. Screened records[5] and records with negative incomes were eliminated[6] from the initial sample of 9,492 leaving a sample of 9,350 families. The sample was further split into those families with male heads and those with female heads. The sample of male heads has 5,371 observations and the sample of female heads has 3,979. The income data we use is the before–tax family income. We calculated the value of the S-Gini AKS index, and the mean of order r AKS index for the full sample and the two subsamples. Results of this estimation for the S-Gini are in Table 1 for the alternative sets of quantiles (10, 20, and 50). Calculations were done with a FORTRAN program that is available from the authors.

[5]Screened records are records which are disguised by Statistics Canada to preserve the anonymity of the survey respondants. Typically they are very large families or have some other similar identifying characteristic.

[6]Negative incomes need to be eliminated from the sample because the mean of order r EDE function is not defined when income is negative.

	Traditional	Number of Quantiles					
	Estimate	10		20		50	
δ	$I(y)$	$\hat{I}(y)$	SE	$\hat{I}(y)$	SE	$\hat{I}(y)$	SE
All Families							
1.5	.2354	.2028	.0017	.2237	.0018	.2324	.0019
2.0	.3624	.3502	.0024	.3593	.0024	.3620	.0025
5.0	.6265	.6202	.0028	.6249	.0028	.6263	.0028
Male Heads							
1.5	.2148	.1824	.0021	.2032	.0024	.2119	.0023
2.0	.3327	.3206	.0031	.3296	.0031	.3323	.0031
5.0	.5900	.5828	.0041	.5882	.0041	.5898	.0041
Female Heads							
1.5	.2563	.2229	.0028	.2442	.0029	.2531	.0030
2.0	.3890	.3763	.0037	.3857	.0038	.3884	.0038
5.0	.6417	.6362	.0039	.6404	.0040	.6415	.0039

Table 1: Estimates of S-Gini AKS index

Comparing columns 2, 3, 5, and 7 gives an indication of the accuracy of the approximations of the index. Column 2 shows the value of the index when calculated the traditional way using the ψ function on each observation individually. The values in columns 3, 5 and 7 show the estimates when the number of quantile groups in equation (23), $S + 1$, is 10, 20, and 50 respectively. The quantile approximation tends to be very close to the more traditional estimate, even in cases where the degree of inequality aversion, δ, is very high, for as few as 20 groups, and very close indeed at 50 quantile groups. It is thus recommended that, preferably, 50 quantile groups be used in applied work, whereas the use of deciles is rather crude. Note also that family income inequality tends to be quite markedly higher for female headed households than male headed households.

The values in columns 4, 6, and 8 of Table 1 are the asymptotic standard

	Number of Quantiles					
	10		20		50	
δ	lower	upper	lower	upper	lower	upper
All Families						
1.5	.1995	.2061	.2202	.2272	.2289	.2359
2.0	.3455	.3549	.3546	.3640	.3571	.3669
5.0	.6147	.6257	.6194	.6304	.6208	.6318
Male Heads						
1.5	.1783	.1865	.1985	.2079	.2074	.2164
2.0	.3145	.3267	.3235	.3357	.3262	.3384
5.0	.5748	.5908	.5802	.5962	.5818	.5978
Female Heads						
1.5	.2174	.2284	.2385	.2499	.2472	.2590
2.0	.3690	.3836	.3783	.3931	.3810	.3958
5.0	.6286	.6438	.6326	.6482	.6339	.6491

Table 2: Confidence Intervals for S–Gini AKS Index

errors of the corresponding approximation estimates of the AKS inequality indices. These standard errors indicate that the estimates are very precise. The standard errors can be used to calculate 95% asymptotic confidence intervals for the values of the indices[7]. These are given in Table 2.

The confidence intervals confirm the reaction that the degree of inequality is estimated very precisely. In most cases, the size of the interval is only about .01. An interesting result is that the confidence intervals for the male and female headed households do not overlap. This suggests that there is a significant difference between male and female heads and that the observed difference in the estimates is not a result of sampling error. Table 3 presents sample t-statistics for the null hypothesis that the underlying inequality value is the same for the male and female heads. As can be seen, the smallest t-value

[7]The confidence intervals are calculated as $\hat{I}(y) \pm 1.96se$.

t-statistics			
Number of Quantiles			
δ	10	20	50
1.5	-11.57	-10.89	-10.89
2.0	-11.53	-11.43	-11.43
5.0	-9.43	-9.11	-9.13

Table 3: Test Statistics for Equality of Male and Female Heads, S–Gini

in absolute value is -9.11 which allows rejection of the null at all reasonable sizes.

An intermediate result which is of some independent interest is the quantile estimates of the EDE, $\hat{e}$, the mean income, $\hat{\mu}$, and the covariance matrix between the two. These values are in Table 4 and can be used in distributionally sensitive cost–benefit analysis or to measure the level of economic well–being of society. The covariances are estimated from Theorem 3 above. For the sake of presentational simplicity, only the results for $S+1=10$, and $S+1=50$ are presented. One item of note from Table 4 is that male–headed households have significantly higher welfare than female–headed households. The differences in well–being between the male–headed and female–headed families is much more stark than is suggested by the inequality measures by themselves. On average, male headed households have well–being a third again larger than that of female headed households. In addition, note that the approximated EDE is estimated more precisely, the higher is the degree of inequality aversion, and the fewer groupings that are used in the estimation.

The estimation can be repeated for the mean of order r inequality index. These results are presented in Tables 5 and 6. The approximation results for the mean of order r are not quite as accurate as for the S–Gini. It is still

S–Gini						
δ	$\hat{e}$	$\hat{\mu}$	Var($\hat{e}$)	Var($\hat{\mu}$)	Cov($\hat{e}, \hat{\mu}$)	Corr($\hat{e}, \hat{\mu}$)
10 Quantile Groups						
All Families						
1.5	35229.98	44190.76	63638.8	99635.8	76016.5	.9546
2.0	28715.90	44190.76	49741.4	99635.8	62022.6	.8810
5.0	16785.49	44190.76	26582.8	99635.8	33406.8	.6491
Male Heads						
1.5	40276.41	49263.96	119607.9	178193.1	139163.6	.9532
2.0	33470.53	49263.96	98764.1	178193.1	116457.7	.8779
5.0	20552.15	49253.96	69734.7	178193.1	71686.1	.6431
Female Heads						
1.5	29018.71	37342.82	114972.9	205004.4	146660.2	.9553
2.0	23289.69	37342.82	80582.8	205004.4	112904.9	.8784
5.0	13584.77	37342.82	30908.8	205004.4	49494.6	.6218
50 Quantile Groups						
All Families						
1.5	33915.66	44184.85	61977.3	99635.8	74034.7	.9421
2.0	28190.23	44184.85	49216.6	99635.8	61057.0	.8719
5.0	16511.80	44184.85	26341.6	99635.8	33039.4	.6449
Male Heads						
1.5	38821.22	49261.45	115541.9	178193.1	135589.6	.9450
2.0	32982.13	49261.45	97003.8	178193.1	114701.6	.8724
5.0	20207.68	49261.45	69210.6	178193.1	71110.9	.6403
Female Heads						
1.5	27883.61	37332.56	108464.3	205004.4	140958.5	.9453
2.0	22831.13	37332.56	77850.0	205004.4	110032.1	.8710
5.0	13384.56	37332.56	30228.5	205004.4	48631.7	.6178

Table 4: Estimates of S-Gini EDE and Mean Income

	Traditional	Number of Quantiles					
	Estimate	10		20		50	
r	$I(y)$	$\hat{I}(y)$	SE	$\hat{I}(y)$	SE	$\hat{I}(y)$	SE
All Families							
0.5	.1065	.1029	.0013	.1050	.0033	.1060	.0039
-0.5	.3157	.2996	.0014	.3057	.0034	.3090	.0040
-1.0	.4260	.3836	.0014	.3927	.0034	.3984	.0042
Male Heads							
0.5	.0909	.0873	.0016	.0894	.0045	.0905	.0057
-0.5	.2831	.2606	.0016	.2695	.0046	.2743	.0060
-1.0	.4031	.3405	.0016	.3546	.0047	.3639	.0063
Female Heads							
0.5	.1201	.1161	.0022	.1184	.0049	.1194	.0055
-0.5	.3318	.3196	.0023	.3246	.0050	.3267	.0056
-1.0	.4270	.3984	.0023	.4052	.0050	.4085	.0057

Table 5: Estimates of Mean of Order r AKS index

t-statistics Number of Quantiles			
r	10	20	50
0.5	-10.58	-4.35	-3.64
-0.5	-21.05	-8.11	-6.38
-1.0	-20.66	-7.37	-5.24

Table 6: Test Statistics for Equality of Male and Female Heads, Mean of Order r

quite reasonable, however, when estimated using fifty quantile groupings. The difficulty in the estimation when compared to the S–Gini is a result of the sensitivity of the two indices to inequality in different locations on the income distribution. The S–Gini is affected most by inequality close to the mean of a distribution, while the mean of order r is affected mostly by inequality in the tails of the distribution[8]. Therefore our approximation where the quantile means are used will affect the mean of order r rather more than the S-Gini. With the size of current microdata sets, working with fifty partitions of the income distribution is very easily handled.

Qualitatively the results for the mean of order r indices are similar to the results for the S–Gini. The size of the t-statistics for the test of identical inequality for the male and female headed subsamples is not quite as striking as with the S–Gini, but the conclusions of significant difference does not change (see Table 6).

6 Conclusion

This paper presents a method of estimation and a derivation of the asymptotic distribution of two general classes of inequality indices, the AKS relative inequality measures and the KBD absolute inequality measures. The approximation method used is applicable to a very large number of possible social evaluation functions, is quite straightforward to compute, and provides distribution–free sampling inferences. The examples in the last section demonstrate the ease and usefulness of the technique. The approach can clearly be used for other inequality measures as well.

[8]See Blackorby and Donaldson (1978).

References

[1] Barrett, G. (1993): 'The Asymptotic Distribution of the S-Gini Indices of Inequality: An Application to Male and Female Hourly Earnings Differences,' Discussion Paper 93-35, Department of Economics, University of British Columbia.

[2] Beach, C. and R. Davidson (1983): 'Distribution-Free Statistical Inference with Lorenz Curves and Income Shares,' *Review of Economic Studies*, 50, 723–735.

[3] Beach, C., J. Formby, K.V. Chow, and G. Slotsve (1992): 'Statistical Inference for Decile Means,' *Economics Letters*, 45, 161–167.

[4] Beach, C. and J. Richmond (1985): 'Joint Confidence Intervals for Income Shares and Lorenz Curves,' *International Economic Review*, 26, 439–450.

[5] Blackorby, C. and D. Donaldson (1978): 'Measures of Relative Inequality and Their Meaning in Terms of Social Welfare,' *Journal of Economic Theory*, 18, 59–80.

[6] Chakravarty, S. (1990): *Ethical Social Index Numbers*, Springer-Verlag, Berlin.

[7] Cowell, F. (1977): *Measuring Inequality*, Oxford University Press.

[8] Cowell, F. (1989): 'Sampling Variance and Decomposable Inequality Measures', *Journal of Econometrics*, 42, 27–41.

[9] Donaldson, D. and J. Weymark (1980): 'A Single-Parameter Generalization of the Gini Indices of Inequality', *Journal of Economic Theory*, 22, 67–86.

[10] Donaldson, D. and J. Weymark (1983): 'Ethically Flexible Gini Indices for Income Distributions in the Continuum,' *Journal of Economic Theory*, 29, 353–358.

[11] Gastwirth, J. (1974): 'Large Sample Theory of Some Measures of Inequality', *Econometrica*, 42, 191–196.

[12] Gastwirth, J. (1975): 'The Estimation of a Family of Measures of Economic Inequality,' *Journal of Econometrics*,3, 61–70.

[13] Layard, R., and A. Zabalza (1979): 'Family Income Distribution: Expalnation and Policy Evaluation,' *Journal of Political Economy*, 87, S133–S161.

[14] Rao, C. (1965): *Linear Statistical Inference and its Applications*, John Wiley and Sons, New York.

[15] Thistle, P. (1990): 'Large Sample Properties of Two Inequality Indices' *Econometrica*, 58, 725–728.