Discussion Paper

Central Bureau of Statistics, P.B. 8131 Dep, 0033 Oslo 1, Norway

No. 14 1 Juni 1986

ON THE PROBLEM OF MEASURING INEQUALITY

BY

ROLF AABERGE

CONTENTS

ABSTRACT

The combination of the Lorenz curve and the Gini coefficient is a widely used tool for measuring inequality in the distribution of income. In the present paper we suggest inequality curves which possess simple interpretations similar to the Lorenz curve and measures of inequality constructed by integral functional mappings of these curves. In addition we introduce inequality decompositions by factor components.

ON THE PROBLEM OF MEASURING INEQUALITY

By

Rolf Aaberge
Central Bureau of Statistics,
Oslo Dep., Oslo, Norway.

1. INTRODUCTION

In economic and sociological litterature, the egalitarian definition of equality in the distribution of a recource, which for convenience we shall refer to as income, is usually applied. The equal income distribution is attained if each unit in the population receives the same income. Inequality is defined as deviation from the state of equality and restricted to satisfy the principles of transfers and scale invariance. The principle of transfers means that inequality is reduced if we transfer income from a richer to a poorer person and the transfer is not so large that the receiver becomes richer than the donor. The scale invariance principle means that inequality will remain unchanged if we increase every unit's income by the same proportion.

The Lorenz curve is a transformation of the cumulative distribution function, which gives a graphical representation of inequality in the distribution function. Within the class of transformations satisfying the principles of transfers and scale invariance, there is a one to one correspondence between the Lorenz curve and the cumulative distribution function. Consequently, the Lorenz curve preserves information about inequality in accordance with the above definition. The Lorenz curve relates the cumulative proportion of income units to the cumulative proportion of income received when units are arranged in ascending order of their income and takes the form of a straight line, the L-line, if and only if all units in the population receive the same income. The L-line represents the equality reference of the Lorenz curve. If any units have unequal incomes the Lorenz curve is a convex function falling below the L-line.

It will be useful to distinguish between the problem of ranking of distributions and the problem of quantifying the differences in inequality between distributions. As a first step, we can use the relation of the Lorenz curve of one distribution being strictly inside that of another as a criterion of ranking of distributions. However, since Lorenz curves may intersect, the criterion of Lorenz curve ranking is incomplete.

In an attempt to establish a general ranking principle several authors have tried to derive criteria from a welfare theoretic approach (c.f. Dalton (1920) and Atkinson (1970)). This approach is, however, unsatisfactory since it requires a complete specification of the welfare functions and these specifications are neither testable nor justified from theoretical arguments.

Besides giving an excellent survey of the literature on measures of inequality, Nygård and Sandström (1981, pp. 122-131) discuss the problem concerning the selection of a welfare function and they conclude that the welfare theory does not contribute to the solution of the ranking problem.

An alternative approach is to construct mappings (functionals) of the Lorenz curve into the real line. The conventional measures of inequality which satisfy the principles of transfers and scale invariance are in fact functional mappings of the Lorenz curve. However, these mappings are not expressed in closed forms. A well known exception is the Gini coefficient, which is equal to twice the area between the L-line and the Lorenz curve. In other words, the Gini coefficient is a mapping of the Lorenz curve into the real line. We shall denote this mapping the integral functional. The Gini coefficient gives a strategy for both ranking distributions and quantifying the differences in inequality between distributions. On the other hand, this strategy must necessarily suffer from certain inconveniences. Evidently, no single measure can reflect all aspects of inequality of an income distribution, only summarize it to a certain extent. Consequently, it will be important to have alternatives to the Gini coefficient. We may for instance derive a family of competitors to the Gini coefficient by forming alternative mappings of the Lorenz curve into the real line. Such an approach will, however, not preserve the attractive geometric interpretation given by the integral functional. An alternative approach is to construct one to one transformations of the Lorenz curve and derive a family of inequality measures by making integral functional mappings of these transformations. One to one transformations of the Lorenz curve fullfilling the principle of transfers will be called inequality curves.

The advantage of this approach is that the curves and measures of inequality can be expressed as functions of the Lorenz curve, and consequently, important results of economic applications of the Lorenz curve ((Jakobsson (1976), Kakwani (1977)) can easily be translated into these families of curves and measures of inequality.

In this paper we propose inequality curves which possess simple interpretations similar to the Lorenz curve and measures of inequality constructed by integral functional mappings of these curves. The actual measures will not necessarily rank distributions in the same order.

Each of them will, however, due to their attractive geometric interpretation, provide detailed information of the actual ranking. This can be done by the plotting of corresponding inequality curves. In the present paper we will discuss the properties of the proposed measures and show how they reflect various aspects of inequality of an income-distribution. For practical purposes it usually will be suitable to apply two or several of these measures rather than a single one.

Section 2 introduces the curves of inequality and section 3 the corresponding measures of inequality. Section 4 gives methods of decomposition. Section 5 deals with the estimation of the introduced curves and measures of inequality.

2. CURVES OF INEQUALITY

Let X be an income variable with cumulative distribution function $F(\cdot)$ and mean $EX = \mu$. Let $[0,\infty\rangle$ be the domain of F where $F^{-1}(0) = 0$. The Lorenz curve $L(\cdot)$ for F is defined by (Gastwirth (1971))

$$L(u) = \frac{1}{\mu}\int_0^u F^{-1}(t)dt, \quad 0 \leq u \leq 1, \tag{2.1}$$

where $F^{-1}(t) = \inf\{x: F(x) \geq t\}$ is the left inverse of F.

As pointed out in the introduction the Lorenz curve gives a graphical representation of an approved definition of the consept of inequality. The family $\mathcal{T}$ of one to one transformations of the Lorenz curve fullfilling the principle of transfers obviously provides alternative expressions of this specific representation. The members of the family $\mathcal{T}$ will be called inequality curves. In this section we will establish some inequality curves which have attractive economic interpretations.

Let the curves $M(\cdot)$, $N(\cdot)$, $P(\cdot)$ and $Q(\cdot)$ be defined by

$$M(u) = \frac{L(u)}{u} \tag{2.2}$$

$$N(u) = \frac{L(u)}{1-L(1-u)} \tag{2.3}$$

$$P(u) = \left(\frac{1-u}{1-L(u)}\right)\left(\frac{L(u)}{u}\right) \tag{2.4}$$

and

$$Q(u) = \frac{1-L(u)}{1-u} \tag{2.5}$$

Note that the above curves have the following limiting properties

$$\lim_{u\to 0} M(u) = 0,\ \lim_{u\to 0} N(u) = 0,\ \lim_{u\to 0} P(u) = 0,\ \lim_{u\to 1} P(u) = \frac{\mu}{F^{-1}(1)} \quad \text{and}$$

$$\lim_{u\to 1} Q(u) = \frac{F^{-1}(1)}{\mu}$$

The above definitions show that there are one to one correspondences between $L(\cdot)$ and $M(\cdot)$, $N(\cdot)$, $P(\cdot)$ and $Q(\cdot)$, respectively. In addition we see that $M(\cdot)$, $N(\cdot)$, $P(\cdot)$ and $Q(\cdot)$ satisfy the principle of transfers, and are thus members of the family $\mathcal{F}$.

By introducing the conditional mean functions $H(\cdot)$ and $K(\cdot)$ defined by

$$H(u) = E[X \quad X\leq F^{-1}(u)] = \frac{1}{u}\int_0^u F^{-1}(t)\,dt, \qquad 0\leq u\leq 1, \tag{2.6}$$

and

$$K(u) = E[X \quad X\geq F^{-1}(u)] = \frac{1}{1-u}\left[\mu - \int_0^u F^{-1}(t)\,dt\right], \quad 0\leq u\leq 1, \tag{2.7}$$

respectively, we get from (2.1), (2.2), (2.3), (2.4) and (2.5) that the inequality curves M, N, P and Q can be written on the following forms

$$M(u) = \frac{E(X|X\leq F^{-1}(u))}{EX}, \quad 0\leq u\leq 1, \tag{2.8}$$

$$N(u) = \frac{E(X|X\leq F^{-1}(u))}{E(X|X\geq F^{-1}(1-u))}, \quad 0\leq u\leq 1, \tag{2.9}$$

$$P(u) = \frac{E(X|X\leq F^{-1}(u))}{E(X|X\geq F^{-1}(u))}, \quad 0\leq u\leq 1, \tag{2.10}$$

and

$$Q(u) = \frac{E(X|X\geq F^{-1}(u))}{EX}, \quad 0\leq u\leq 1, \tag{2.11}$$

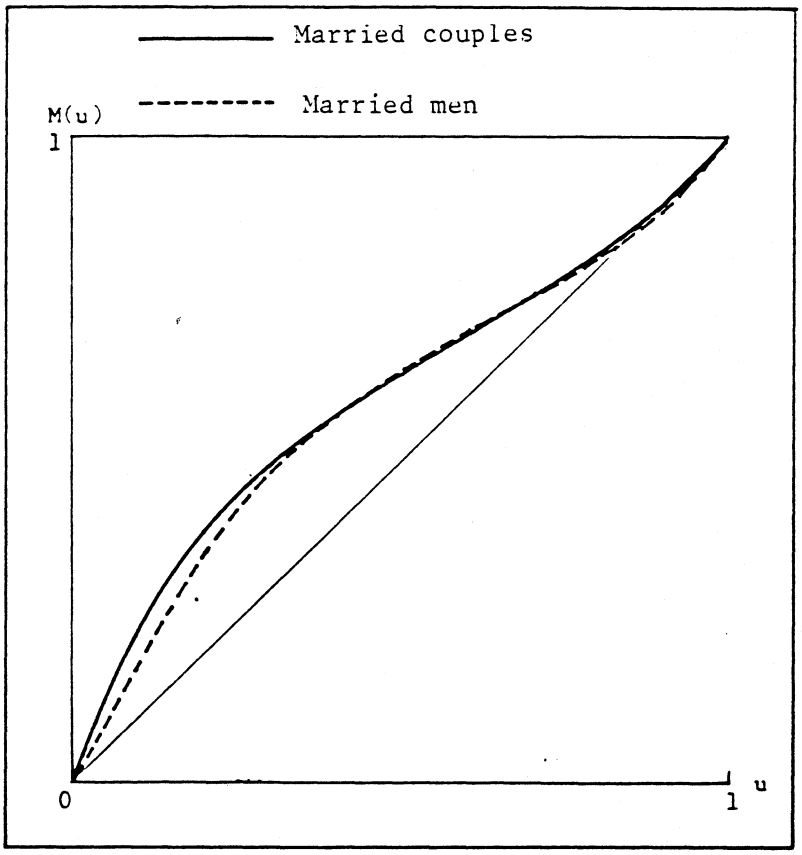
These expressions show that the inequality curves M, N, P and Q possess simple interpretations similar to the Lorenz curve.

For a fixed u, M(u) expresses the ratio between the mean income of the poorest 100 u percent of the population and the mean income of the population, N(u) is the ratio between the mean income of the poorest 100 u percent and the mean income of the richest 100 u percent of the population, P(u) is the ratio between the mean income of the poorest 100 u percent and the mean income of the richest 100 (1-u) percent of the population, and Q(u) is the ratio between the mean income of the richest 100 (1-u) percent of the population and the mean income of the population.

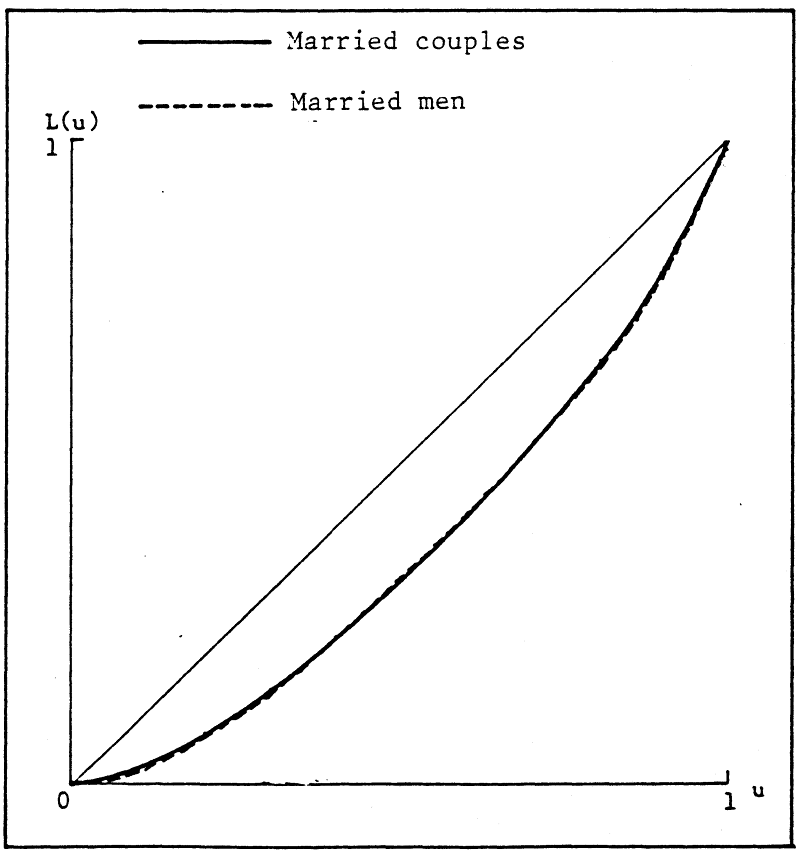
As mentioned above, the straight line joining the points (0,0) and (1,1) is called the egalitarian line of the Lorenz curve, which means that each unit receives the same income. Thus, equations (2.2), (2.3), (2.4) and (2.5) imply that the egalitarian lines of M, N, P and Q coincide with the horizontale line joining the points (0,1) and (1,1).

Note that the universes of M-curves, N-curves and P-curves are each bounded by a unit square. Therefore visually, there is a sharper distinction between two different M-curves, N-curves and P-curves, respectively, than between the two corresponding Lorenz-curves. As an illustration, we have plotted the M-curves (figure 1) and the Lorenz-curves (figure 2) of the income distributions of Norwegian married couples and Norwegian married men in 1979.

Figur 1. M-curves



Figur 2. Lorenz-curves



As can be seen from figures 1 and 2, the plots of the M-curves show that there is larger inequality in the lower tail of the income distribution of married men than in the lower tail of the income distribution of married couples, while the corresponding Lorenz-curves apparently do not display the same information. Our example demonstrates that there may be differences in inequality between the lower tails of two distribution functions, which the plots of the corresponding Lorenz-curves fail to detect.

Note that the M-curve will be the diagonal line joining the points (0,0) and (1,1) if and only if the underlying distribution is uniform (0,a). Thus, the M-curve for a uniform (0,a) distribution represents an interesting additional reference line.

Since the lower and central parts of the M-curves in figure 1 are lying above the diagonal line and the upper parts of the M-curves are lying below the diagonal line, we can state that there is less inequality in the lower and central parts and larger inequality in the upper parts of the current distributions compared to the inequality possessed by a uniform (0,a) distribution.

In table 1 we present inequality curves generated by four well-known distribution functions.

Table 1. The Form of L, M, N, P and Q for some Distributions Functions

Inequality Curve	Cumulative Distribution Function			
	$F(x)=\begin{cases}1, & x\geq a\\ 0, & x<a\end{cases}$	$F(x)=1-e^{-ax}$ $x>0, a>0$	$F(x)=\frac{x}{a}$ $0\leq x\leq a$	$F(x)=1-(\frac{a}{x})^{a}$ $x>a,\ a>1$
$L(u)$	u	$u+(1-u)\log(1-u)$	u^2	$1-(1-u)^{(\frac{a-1}{a})}$
$M(u)$	1	$1+(\frac{1-u}{u})\log(1-u)$	u	$\frac{1-(1-u)^{(\frac{a-1}{a})}}{u}$
$N(u)$	1	$\frac{u+(1-u)\log(1-u)}{u(1-\log u)}$	$\frac{u}{2-u}$	$u^{(\frac{1-a}{a})}-(\frac{1-u}{u})^{(\frac{a-1}{a})}$
$P(u)$	1	$\frac{u+(1-u)\log(1-u)}{u(1-\log(1-u))}$	$\frac{u}{1+u}$	$\frac{(1-u)^{a^{-1}}-(1-u)}{u}$
$Q(u)$	1	$1-\log(1-u)$	$1+u$	$(1-u)^{-a^{-1}}$

3. MEASURES OF INEQUALITY

The concept of inequality, as noted earlier, is multidimensional. Therefore, it is necessary to apply several measures in order to characterize various aspects of inequality in a distribution of income.

A familar approach in deriving measures of inequality is to transform measures of dispersion into measures of inequality. This type of measures is, however, difficult to relate to the Lorenz formalization of inequality. A more attractive approach is to derive measures of inequality from the family of inequality curves ($\mathcal{T}$) by applying the integral functional. The resulting family of measures of inequality ($\mathcal{I}$) is defined by

$$I = k_1 + k_2 \int_0^1 T(u)du \tag{3.1}$$

where $T \in \mathcal{T}$ and k_1 and k_2 are suitable normalizing constants.

All members of the family $\mathcal{I}$ satisfy the principles of transfers and scale invariance. This fact is an immediate consequence of the properties for $\mathcal{T}$ and the integral functional.

The Gini coefficient G, defined by

$$G = 2 \int_0^1 (u-L(u))du \ , \tag{3.2}$$

is a member of the family $\mathcal{I}$.

The specific inequality curves introduced in section 2 result in the following members of I,

$$A = \int_0^1 (1-M(u))du \ , \tag{3.3}$$

$$B = \int_0^1 (1-N(u))du \ , \tag{3.4}$$

$$C = \int_0^1 (1-P(u))du \tag{3.5}$$

and

$$D = \int_0^1 (Q(u) - 1)du \ . \tag{3.6}$$

A, B, C and D are defined by the area between the inequality curves $M(\cdot)$, $N(\cdot)$, $P(\cdot)$ and $Q(\cdot)$ respectively, and the corresponding egalitarian lines, which in these cases are the horizontal line between the points (0,1) and (1,1). The range for A, B and C is [0,1], while the range for D is $[0,\infty\rangle$. A [0,1] normalized version of D is given by

$$D' = \frac{D}{(D^2+1)^{\frac{1}{2}}} \ . \tag{3.7}$$

A desirable property of an inequality measure is that it should equal zero when the underlying distribution function expresses perfect equality. The Gini coefficient and each of the above introduced measures of inequality possess this specific property. Another desirable property of an equality measure is that the maximum attainable value is one, i.e. the value one is obtained if one unit receives all incomes and the others zero income. The measures A, B, C, D' and G possess this property.

Also note that A, B, C, D' and G take the values •500, •614, •693, •447 and •333, respectively, if the underlying distribution function is uniform (0,a).

Alternative expressions for A, B, C and D, indicating probabilistic interpretations, are given by

$$A = \frac{E[X-E(X|X\leq Y)]}{EX} \tag{3.8}$$

where X and Y are independent and identically distributed

$$B = E[\frac{E(X|X\geq F^{-1}(1-U)) - E(X|X\leq F^{-1}(U))}{E(X|X\geq F^{-1}(1-U))} \tag{3.9}$$

where U is a uniform (0,1) distributed variabel independent of X and F is the distribution function of X ,

$$C = E[\frac{E(X|X\geq Y) - E(X|X\leq Y)}{E(X|X\geq Y)}] \tag{3.10}$$

and

$$D = \frac{E[E(X|X\geq Y) - X]}{EX} \tag{3.11}$$

The advantage of adopting (3.1) with respect to interpretation is obvious. In a simple way each of the derived measures can be discussed in relation to inequality curves that have independent economic interpretation. Each measure reflects the properties of the corresponding inequality curve.

Now we shall examine the derived measures with respect to sensitivity to transfers. As we shall see, there are in fact important differences in sensitivity to transfers at different parts of the underlying distribution. As pointed out by Atkinson (1970), the Gini coefficient attaches more weight to tranfers in the centre of an unimodal distribution than at the tails. By expressing the above measures of inequality in terms of the Lorenz curve it is easily seen that A, B and C attach more weight to transfers at the lower tail than at the centre and at the upper tail. If, therefore, one wants to give more weight to transfers at the lower end of the distribution than at the top, all these measures are appropriate. Note that A weights transfers at the lower tail more heavily than B and C.

D and D' attach more weight to transfers in the upper tail of a distribution than at the centre and the lower tail.

For practical purposes it will usually be appropriate to apply a few of the above discussed measures of inequality simultaneously. In that way we will obtain more detailed but still compact information of inequality in distributions of income.

4. DECOMPOSITION BY FACTOR COMPONENTS

In section 2 we introduced various measures of inequality, designed to summarize or aggregate the inequality of an income distribution function. Judgements about the importance of various influences on the inequality of a distribution function are another aspect of the analysis of inequality. In economic literature it is common to relate these judgments to measures of inequality and to attempt to decompose these measures of inequality into relevant component contributions.

We will study decompositions for situations in which the income of individuals or households is expressed as the sum of incomes from different factor components, such as earnings, investment income, negative and positive transfer payments, etc. The main purpose of deriving this type of decompositions is that they are useful tools for assessing the contributions of different income factors on the inequality of the distribution of total income.

Let Z be a random variable (income variable) with distribution function $F(\cdot)$ and mean μ. We assume that the income variable Z is the sum of incomes from s different factor components Z^k with distribution functions F_k and corresponding inequality measures G_k, A_k, B_k, C_k and D_k, $k=1,2,\ldots,s$.

Now we will decompose the inequality measures A, B, C, D and G according to the s factor components.

Since $Z = \sum_{k=1}^{s} Z^k$ we obtain the following expression of the conditional mean function $H(\cdot)$, defined by (2.6),

$$H(u) = \sum_{k=1}^{s} H_k(u) \tag{4.1}$$

where

$$H_k(u) = E(Z^k \mid Z \leq F^{-1}(u)) .$$

The mean μ is given by

$$\mu = H(1) = \sum_{k=1}^{s} \mu_k \qquad (4.2)$$

where

$$\mu_k = EZ^k .$$

By substituting (4.1) and (4.2) in (3.2), (3.3), (3.4), (3.5) and (3.6), respectively, we get the following decomposition rules

$$G = \sum_{k=1}^{s} u_k(G) = \sum_{k=1}^{s} \frac{\mu_k}{\mu} \frac{\gamma_k}{G_k} G_k \qquad (4.3)$$

where

$$\gamma_k = 1-2 \int_0^1 l_k(u)du \text{ and } l_k(u) = u\frac{E(Z^k | Z \leq F^{-1}(u))}{\mu_k},$$

$$A = \sum_{k=1}^{s} u_k(A) = \sum_{k=1}^{s} \frac{\mu_k}{\mu} \frac{\alpha_k}{A_k} A_k \qquad (4.4)$$

where

$$\alpha_k = 1 - \int_0^1 m_k(u)du \text{ and } m_k(u) = \frac{E(Z^k | Z \leq F^{-1}(u))}{\mu}$$

$$B = \sum_{k=1}^{s} u_k(B) = \sum_{k=1}^{s} \frac{\mu_k}{\mu} \frac{\beta_k}{B_k} B_k \qquad (4.5)$$

where

$$\beta_k = 1 - \int_0^1 t_k(u)du \text{ and } t_k(u) = \frac{\mu E(Z^k | Z \leq F^{-1}(u))}{\mu_k E(Z | Z \geq F^{-1}(1-u))},$$

$$C = \sum_{k=1}^{s} u_k(C) = \sum_{k=1}^{s} \frac{\mu_k}{\mu} \frac{\pi_k}{C_k} C_k \qquad (4.6)$$

where

$$\pi_k = 1 - \int_0^1 p_k(u)du \text{ and } p_k(u) = \frac{\mu E(Z^k|Z \leq F^{-1}(u))}{\mu_k E(Z|Z \geq F^{-1}(u))}$$

and

$$D = \sum_{k=1}^{s} u_k(D) = \sum_{k=1}^{s} \frac{\mu_k}{\mu} \frac{\theta_k}{D_k} D_k \tag{4.7}$$

where

$$\theta_k = \int_0^1 q_k(u)du - 1 \text{ and } q_k(u) = \frac{E(Z^k|Z \geq F^{-1}(u))}{\mu_k} .$$

We define the ratios (γ_k/G_k), (α_k/A_k), (β_k/B_k), (π_k/C_k) and (θ_k/D_k) to equal zero when the state of equality for factor k occurs.

A discrete version of (4.3) was proposed by Rao (1969) and later a generalized version corresponding to (4.3) was introduced and discussed by Nygård and Sandstrøm (1981). By applying the decomposition method for the Gini coefficient, Sandstrøm (1982) studies the effect of various income sources on the inequality of total income.

The present decomposition rules are all on the same form. It is therefore sufficient to present a detailed discussion for only one of the derived decomposition rules. We shall relate this discussion to the decomposition rule for A.

If A is the inequality measure chosen, we see from (4.4) that $u_k(A)$ is the inequality contribution of factor k. The term (μ_k/μ) expresses the income share of factor k and A_k is the A-inequality of the distribution of Z^k. The term α_k is an expression of interaction between Z^k and Z and will be called the interaction component of factor k. The interaction component α_k is, loosely spoken, the conditional A-inequality of factor k given by the units rank order in total income. By studying the sign of α_k we see whether or not factor k acts equalizing or disequalizing. A positive value on α_k would increase the inequality and a negative value would decrease the inequality when $\mu_k>0$. If $\mu_k<0$, we have the opposite result.

There are two characters of factor k which have influence on the value of α_k. The first character is the A-inequality (A_k) of the distribution of the factor k variable Z^k and the second character is the location of the factor k incomes in the rank order of total income. Therefore, we have expressed the interaction component as a product of A_k and (α_k/A_k). The last term informs about the location effect of factor k on the A-inequality in the distribution of total income (F).

Notice that α_k is determined by the area below the curve $m_k(\cdot)$, henceforth called the interaction curve of factor k. For a fixed u, $m_k(u)$ gives the ratio between the factor k mean of those units having total income less than or equal to $F^{-1}(u)$ and the factor k mean μ_k. If, for instance, Z^k is an earning variable then $m_k(1/2)$ gives the ratio between the mean earning of the lower half of the income distribution F and the mean earning μ_k. In the particular situation where $E(Z^k|Z \leq F^{-1}(u)) = E(Z^k|Z^k \leq F_k^{-1}(u))$ for every u, i.e. the units rank order in earning coincides with their rank order in total income, we have that $m_k(u) = M_k(u)$ for all u and thus $\alpha_k = A_k$, where $M_k(\cdot)$ is the M-curve of the factor k variable.

If $\alpha_k=0$, factor k will have a neutral effect on the A-inequality of the distribution of total income, but still factor k can have different effects on the A-inequality of various parts of this distribution. The effect is neutral in every part of the distribution if and only if $m_k(u) = 1$ for all u. Thus, this shows the importance of plotting the interaction curves instead of solely being concerned about the corresponding interaction components.

As stated earlier, the interpretation of the remaining decomposition rules is similar to the interpretation of the decomposition rule for A. The differences between A, B, C, D and G pointed out in section 3, are reflected in the relations between the corresponding interaction components. We have, for instance, that α_k attach more weight than γ_k to transfers concerning the factor k incomes of units with small total incomes.

The interaction components α_k, β_k, π_k, θ_k and γ_k satisfy

$$-D_k \leqq \alpha_k \leqq A_k\ , \tag{4.8}$$

where D_k and A_k are the D-inequality and the A-inequality in the distribution of the factor k variable, respectively,

$$1-\int_0^1 \frac{Q_k(1-u)}{Q(1-u)}\,du \leqq \beta_k \leqq 1-\int_0^1 \frac{M_k(u)}{Q(1-u)}\,du\ , \tag{4.9}$$

where $Q_k(\cdot)$ and $M_k(\cdot)$ are inequality curves of the factor k variable defined by (2.8) and (2.5), respectively, and $Q(\cdot)$ is the inequality curve of total income defined by (2.8),

$$1-\int_0^1 \frac{Q_k(1-u)}{Q(u)}\,du \leqq \pi_k \leqq 1-\int_0^1 \frac{M_k(u)}{Q(u)}\,du\ , \tag{4.10}$$

$$-A_k \leqq \theta_k \leqq D_k \tag{4.11}$$

and

$$-G_k \leqq \gamma_k \leqq G_k\ , \tag{4.12}$$

where G_k are the Gini coefficient of the distribution of the factor k variable.

The interaction components of factor k attain their maximum values when the rank order of the units in factor k income and total income, respectively, are identical. The minimum values are obtained when the ordering in factor k income is reversed in relation to the ordering in total income.

5. METHODS OF ESTIMATION

In this section we will give nonparametric methods for estimating the curves of inequality, the measures of inequality and the various components of the decompositions.

5.1. Estimation of Curves and Measures of Inequality

Let $X_1, X_2, \ldots, X_n$ be independent random variables with common distribution function F. When the parametric form of F is not known, it is natural to use the empirical distribution function F_n to estimate F and to use

$$\hat{H}(u) = \frac{1}{u} \int_0^u F_n^{-1}(t)\,dt , \quad 0 \leqq u \leqq 1, \tag{5.1}$$

to estimate $H(u)$, where the left inverse of F_n is defined by

$$F_n^{-1}(t) = \inf \{x : F_n(x) \geqq t\} . \tag{5.2}$$

To give a more explicit expression for $\hat{H}(u)$ we introduce the order statistics $X_{(1)} \leqq X_{(2)} \leqq \ldots \leqq X_{(n)}$. Now for $u \in \langle (i-1)/n, i/n]$ we have $F_n^{-1}(u) = X_{(i)}$ and

$$\hat{H}(u) = \frac{1}{nu} \left(\sum_{j=1}^{i-1} X_{(j)} + (nu - i + 1) X_{(i)} \right) . \tag{5.3}$$

For $u = i/n$ (5.3) reduces to the partial sample mean

$$\hat{H}(\tfrac{i}{n}) = \bar{X}_{(i)}, \quad i=1,2,\ldots,n \tag{5.4}$$

where

$$\bar{X}_{(i)} = \frac{1}{i} \sum_{j=1}^{i} X_{(j)}.$$

Now replacing $H(u)$ by $\hat{H}(u)$ in the expression (2.3) for $L(u)$, we get the following estimate for the Lorenz curve L.

$$\hat{L}(u) = u \frac{\hat{H}(u)}{\hat{H}(1)}, \quad 0 \leqq u \leqq 1. \tag{5.5}$$

Similarly, the estimate of M, N, P and Q are obtained from (5.5) and thus given by

$$\hat{M}(u) = \frac{\hat{L}(u)}{u}, \quad 0 \leqq u \leqq 1, \tag{5.6}$$

$$\hat{N}(u) = \frac{\hat{L}(u)}{1-\hat{L}(1-u)}, \quad 0 \leqq u \leqq 1, \tag{5.7}$$

$$\hat{p}(u) = \frac{1-(u-0)}{1-\hat{L}(u-0)} \cdot \frac{\hat{L}(u)}{u}, \quad 0 \leqq u \leqq 1, \tag{5.8}$$

and

$$\hat{Q}(u) = \frac{1-\hat{L}(u-0)}{1-(u-0)}, \quad 0 \leqq u \leqq 1. \tag{5.9}$$

The introduction of (u-0) in (5.8) and (5.9) guarantees that the empirical inequality curves $\hat{P}$ and $\hat{Q}$ will be right continuous.

Now using (5.4) in (5.5), (5.6), (5.7), (5.8) and (5.9) respectively, we get

$$\hat{L}(\tfrac{i}{n}) = \frac{\sum_{j=1}^{i} X_{(j)}}{\sum_{j=1}^{n} X_j}, \; i=1,2,\ldots,n, \tag{5.10}$$

$$\hat{M}(\tfrac{i}{n}) = \frac{\bar{X}_{(i)}}{\bar{X}}, \; i=1,2,\ldots,n, \tag{5.11}$$

where $\bar{X} = \bar{X}_{(n)}$ is the sample mean,

$$\hat{N}(\tfrac{i}{n}) = \frac{\bar{X}_{(i)}}{\bar{X}'_{(i)}}, \; i=1,2,\ldots,n, \tag{5.12}$$

where

$$\bar{X}'_{(i)} = \frac{1}{i} \sum_{j=1}^{i} X_{(n+1-j)},$$

$$\hat{P}(\tfrac{i}{n}) = \frac{\bar{X}_{(i)}}{\bar{X}'_{(n+1-i)}}, \; i=1,2,\ldots,n, \tag{5.13}$$

and

$$\hat{Q}(\tfrac{i}{n}) = \frac{\bar{X}'_{(n+1-i)}}{\bar{X}}, \; i=1,2,\ldots,n. \tag{5.14}$$

Replacing the theoretical inequality curves by the corresponding empirical inequality curves in (3.2), (3.3), (3.4), (3.5) and (3.6), respectively, we arrive at the following estimators of G, A, B, C and D

$$\hat{G} = 2\int_0^1 (u-\hat{L}(u))du = -\frac{2\sum_{i=1}^{n}\sum_{j=1}^{i} X_{(j)}}{(n+1)\sum_{j=1}^{n} X_j}, \tag{5.15}$$

$$\hat{A} = \int_0^1 (1-\hat{M}(u))du = 1-\frac{\sum_{i=1}^{n}\bar{X}_{(i)}}{n\bar{X}}, \tag{5.16}$$

$$\hat{B} = \int_0^1 (1-\hat{N}(u))du = 1-\frac{1}{n}\sum_{i=1}^{n}\frac{\bar{X}_{(i)}}{\bar{X}'_{(i)}}, \tag{5.17}$$

$$\hat{C} = \int_0^1 (1-\hat{P}(u))du = 1-\frac{1}{n}\sum_{i=1}^{n}\frac{\bar{X}_{(i)}}{\bar{X}'_{(n+1-i)}}, \tag{5.18}$$

and

$$\hat{D} = \int_0^1 (\hat{Q}(u)-1)du = \frac{\sum_{i=1}^{n}\bar{X}'_{(n+1-i)}}{n\bar{X}} - 1. \tag{5.19}$$

In the literature $((n+1)/(n-1))\hat{G}$ is commonly used as an estimator of G. $\hat{G}$ and $((n+1)/(n-1))\hat{G}$ are obviously asymptotic equivalent estimators.

5.2. Estimation of the factor components

Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables with common distribution function F. We assume that $Z_i = \sum_{k=1}^{s} Z_i^k$, $i=1,2,\ldots,n$, where Z_i^k is the factor k income of unit i. This leads to the commonly used estimators of μ and μ_k defined by (4.2)

$$\hat{\mu} = \bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i \tag{5.20}$$

and

$$\hat{\mu}_k = \bar{Z}^k = \frac{1}{n} \sum_{i=1}^{n} Z_i^k . \tag{5.21}$$

Now let $(Z_{(i)}, \tilde{Z}_i^1, \tilde{Z}_i^2, \ldots, \tilde{Z}_i^s)$, $i=1,2,\ldots,n$ be the random vectors $(Z_i, Z_i^1, Z_i^2, \ldots, Z_i^s)$, $i=1,2,\ldots,n$ ordered according to $Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(n)}$. This leads to the empirical conditional mean function

$$\hat{H}_k\left(\frac{i}{n}\right) = \frac{1}{i} \sum_{j=1}^{i} \tilde{Z}_j^k, \quad i=1,2,\ldots,n, \tag{5.22}$$

as an estimate of the conditional mean function $H_k(\frac{i}{n})$ defined by (4.1). Replacing H_k by $\hat{H}_k$ in the current interaction curves and components defined in section 4.1, we obtain the following estimates

$$\hat{l}_k\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^{i} \tilde{Z}_j^k}{\sum_{j=1}^{n} Z_j^k}, \quad i=1,2,\ldots,n;\ k=1,2,\ldots,s, \tag{5.23}$$

$$\hat{\gamma}_k = 1 - \frac{2}{n+1} \sum_{i=1}^{n} \hat{l}_k \left(\frac{i}{n}\right), \; k=1,2,\ldots,s, \tag{5.24}$$

$$\hat{m}_k\left(\frac{i}{n}\right) = \frac{\frac{1}{i} \sum_{j=1}^{i} \tilde{Z}_j^k}{\bar{Z}^k}, \; i=1,2,\ldots,n; \; k=1,2,\ldots,s, \tag{5.25}$$

$$\hat{\alpha}_k = 1 - \frac{1}{n} \sum_{i=1}^{n} \hat{m}_k\left(\frac{i}{n}\right), \; k=1,2,\ldots,s, \tag{5.26}$$

$$\hat{t}_k\left(\frac{i}{n}\right) = \left(\frac{\frac{1}{i} \sum_{j=1}^{i} \tilde{Z}_j^k}{\bar{Z}^k}\right) \left(\frac{\bar{Z}}{\frac{1}{i} \sum_{j=n-i+1}^{n} Z_{(j)}}\right), \; i=1,2,\ldots,n; \; k=1,2,\ldots,s, \tag{5.27}$$

$$\hat{\beta}_k = 1 - \frac{1}{n} \sum_{i=1}^{n} \hat{t}_k\left(\frac{i}{n}\right), \; k=1,2,\ldots,s, \tag{5.28}$$

$$\hat{p}_k\left(\frac{i}{n}\right) = \left(\frac{\frac{1}{i} \sum_{j=1}^{i} \tilde{Z}_j^k}{\bar{Z}^k}\right) \left(\frac{\bar{Z}}{\frac{1}{n-i+1} \sum_{j=i}^{n} Z_{(j)}}\right), \; i=1,2,\ldots,n; \; k=1,2,\ldots,s, \tag{5.29}$$

$$\hat{\pi}_k = 1 - \frac{1}{n} \sum_{i=1}^{n} \hat{p}_k\left(\frac{i}{n}\right), \; k=1,2,\ldots,s, \tag{5.30}$$

$$\hat{q}_k\left(\frac{i}{n}\right) = \frac{\frac{1}{n-i+1} \sum_{j=i}^{n} \tilde{Z}_j^k}{\bar{Z}^k}, \; i=1,2,\ldots,n; \; k=1,2,\ldots,s, \tag{5.31}$$

and

$$\hat{\theta}_k = \frac{1}{n} \sum_{i=1}^{n} \hat{q}_k\left(\frac{i}{n}\right) - 1, \; k=1,2,\ldots,s, \tag{5.32}$$

The estimators for G_k, A_k, B_k, C_k and D_k, based on the sample $Z_1^k, Z_2^k, \ldots, Z_n^k$, are given by (5.15), (5.16), (5.17), (5.18) and (5.19), respectively. The obtained estimator for G given by

$$\hat{G} = \sum_{k=1}^{s} \hat{u}_k(G) = \sum_{k=1}^{s} \frac{\hat{\mu}_k}{\hat{\mu}} \hat{\gamma}_k \tag{5.33}$$

coincides with (5.15).

The results for the estimators A, B, C and D are analogous.

REFERENCES

ATKINSON, A.B.: "On the Measurement of Inequality," Journal of Economic Theory, 2, 244-263, 1970.

DALTON, H.: "The measurement of the Inequality of Incomes," Economic Journal, Vol. 30, 1920.

GASTWIRTH, J.L.: "A General Definition of the Lorenz Curve", Econometrica, 39, 1037-1039, 1971.

JAKOBSSON, U.: "On the Measurement of the Degree of Progression", Journal of Public Economics, 5, 161-168, 1976.

KAKWANI, N.C.: "Applications of Lorenz Curves in Economic Analysis," Econometrica, 45, 719.727, 1977.

NYGÅRD, F., and A. SANDSTRÖM: "Measuring Income Inequality". Almqvist and Wiksell International, Stockholm 1981.

RAO, V.M.: "Two Decompositions of Consentration Ratio," Journal of the Royal Statistical Society, 132, 418-425, 1969.

SANDSTRÖM, A.: "Various Income Sources and Inequality of Total Income in Sweden in 1972 and 1976," Statistical Review, Statistics Sweden, 4, 288-299, 1982.

ISSUED IN THE SERIES DISCUSSION PAPER

No. 1 I. Aslaksen and O. Bjerkholt: Certainty Equivalence Procedures in the Macroeconomic Planning of an Oil Economy.

No. 3 E. Biørn: On the Prediction of Population Totals from Sample Surveys Based on Rotating Panels.

No. 4 P. Frenger: A Short Run Dynamic Equilibrium Model of the Norwegian Production Sectors.

No. 5 I. Aslaksen and O. Bjerkholt: Certainty Equivalence Procedures in Decision-Making under Uncertainty: an Empirical Application.

No. 6 E. Biørn: Depreciation Profiles and the User Cost of Capital.

No. 7 P. Frenger: A Directional Shadow Elasticity of Substitution.

No. 8 S. Longva, L. Lorentsen, and Ø. Olsen: The Multi-Sectoral Model MSG-4, Formal Structure and Empirical Characteristics.

No. 9 J. Fagerberg and G. Sollie: The Method of Constant Market Shares Revisited.

No. 10 E. Biørn: Specification of Consumer Demand Models with Stochstic Elements in the Utility Function and the First Order Conditions.

No. 11 E. Biørn, E. Holmøy, and Ø. Olsen: Gross and Net Capital, Productivity, and the Form of the Survival Function - Some Norwegian Evidence.

No. 12 J. K. Dagsvik: Markov Chains Generated by Maximizing Components of Multidimensional Extremal Processes.

No. 13 E. Biørn, M. Jensen, and M. Reymert: KVARTS - A Quarterly Model of the Norwegian Economy.

No. 14 R. Aaberge: On the problem of measuring inequality.