METRON - *International Journal of Statistics*
2005, vol. LXIII, n. 3, pp. 493-503

ROLF AABERGE - STEINAR BJERVE – KJELL DOKSUM

Decomposition of rank-dependent measures of inequality by subgroups

Summary - The purpose of additive subgroup decomposition is to study the relationship between overall inequality and inequality within and between population subgroups defined by variables like gender, age, education and region of residence. As opposed to the inequality measures that are additively decomposable, the so-called generalized entropy family of inequality measures, the Gini coefficient does not admit decomposition into within- and between-group components but does also require an interaction (overlapping) term. The purpose of this paper is to introduce an alternative decomposition method that can be considered to be a parallel to Lerman and Yitzhaki's (1985) elasticity approach for decomposing the Gini coefficient by income sources, which means that the elasticity of the Gini coefficient with respect to various income components is treated as the basic quantities of the decomposition method. Thus, rather than decomposing the Gini coefficient or any other inequality measure into a within-inequality term, a between-inequality term and eventually an interaction term, the basic quantities of the introduced method are the effects of marginal changes in variables that are used to specify the population subgroups.

Key Words - The Gini coefficient; The Bonferroni coefficient; Rank-dependent measures of inequality; Decomposition by subgroups.

1. Introduction

The most widely used measure of income inequality is the Gini coefficient, which is defined equal to twice the area between the Lorenz curve and its equality reference([1]). The simple and direct relationship between the Gini coefficient and the Lorenz curve appears to be the major reason for its popularity in applied work. However, since empirical analyses of income inequality normally deals with issues that require use of decomposition methods numerous proposals on how to decompose the Gini coefficient by income sources as

([1]) See Giorgi (1990) for a bibliographical portrait of the Gini coefficient.

Received September 2005.

well as by subgroups has occurred in the literature([2]). The purpose of subgroup decomposition is to study the relationship between overall inequality and inequality within and between population subgroups defined by variables like gender, age, education and region of residence([3]). As opposed to the inequality measures that are additively decomposable, the so-called generalized entropy family of inequality measures, the Gini coefficient does not admit decomposition into within- and between-group components. However, by adding an extra term that captures the overlap between the marginal income distributions of subgroups it can be demonstrated that the Gini coefficient can be decomposed into three terms, the within-group term, the between-group term and an interaction term([4]). Note that the interaction term vanishes when there is no overlapping of income ranks between income units belonging to different subgroups; *i.e.* when the income distributions of subgroups do not overlap. However, a number of alternative approaches for decomposing the Gini coefficient and other measures of inequality by subgroups could be defined, see Shorrocks (1984).

The purpose of this paper is to introduce a new method that can be considered to be a parallel to Lerman and Yitzhaki's (1985) elasticity approach for decomposing the Gini coefficient by income sources, which means that the elasticity of the Gini coefficient with respect to various income components is treated as the basic quantities of the decomposition method. Thus, we turn the focus from decomposing the Gini coefficient or any other inequality measure into a within-inequality term, a between-inequality term and eventually an interaction (overlapping) term to the effects of marginal changes in the variables that are used to specify the population subgroups.

2. Decomposition of Lorenz curves and rank-dependent measures of inequality

Let Y be a positive, continuous random variable representing wage or income, $\mathbf{X}$ a random covariate vector. Leaving out the influence of $\mathbf{X}$, the overall Lorenz curve is

$$L(u) = \frac{1}{\mu}\int_0^1 I[t \leq u]F^{-1}(t)dt = \frac{1}{\mu}\int_0^\infty I[y \leq F^{-1}(u)]ydF(y)$$

where $\mu = \int_0^\infty ydF(y)$ is the mean of Y, F^{-1} denotes the left inverse of the distribution function F of Y, and I is the indicator function. $L(u)$ gives the

([2]) See *e.g.* Rao (1969), Kakwani (1977, 1980), Lerman and Yitzhaki (1985), Chakravarty (1990) and Silber (1993) for useful discussions on decomposing the Gini coefficient by income sources.

([3]) See *e.g.* Shorrocks (1984).

([4]) More on the derivation and interpretation of the subgroup decomposition of the Gini coefficient, see Bhatacharya and Mahalanonis (1967), Piesch (1975), Silber (1989), Yitzhaki (1994), Yitzhaki and Lerman (1991), Lambert and Aronson (1993) and Dagum (1997).

proportion of the total amount of income that is owned by the $100u$ poorest percent of the population. We extend this definition to include the influence of covariates by considering the proportion of the total amount of income that is owned by the subpopulation with covariate values $\mathbf{x}$ and with income below the u-th quantile in the entire population. To this end we define the *pseudo-Lorenz regression curve* as

$$\begin{aligned}\Lambda(u|\mathbf{x}) &= \frac{1}{\mu} E\{YI[Y \leq F^{-1}(u)]|\mathbf{X}=\mathbf{x}\} = \frac{1}{\mu}\int_0^{\infty} yI[y \leq F^{-1}(u)]dF(y|\mathbf{x}) \\ &= \int y\left\{I[y \leq F^{-1}(u)]\Big/\int_0^1 F^{-1}(u)du\right\}dF(y|\mathbf{x}), \quad 0 \leq u \leq 1\,. \end{aligned} \tag{2.1}$$

where $F(y|\mathbf{x})$ denotes the distribution function of Y given $\mathbf{X} = \mathbf{x}$. Although this curve differs from the standard Lorenz curve it has the nice property that it is a decomposition of the Lorenz curve in the sense that its expected value equals the Lorenz curve for the total population, *i.e.* by using the iterated expectation theorem, see Bickel and Doksum (2001), we find

$$E\left[\Lambda(u|\mathbf{X})\right] = L(u)\,. \tag{2.2}$$

As (2.1) shows, this definition of Lorenz regression aggregates incomes from the subgroup with covariate vector $\mathbf{x}$, but uses $F^{-1}(u)$ as a *common* reference when computing proportions([5]). This reference quantile $F^{-1}(u)$ is the u-th quantile of the overall income distribution $F(y)$ which is obtained by averaging out $\mathbf{x}$, that is, $F(y) = E[F(y|\mathbf{X})]$. If $\mathbf{X}$ is used to partition the sample space into distinct categories $C_1, \ldots, C_s$ with probabilities $P(C_j) = P(\mathbf{X} \in C_j)$, $j = 1, \ldots, s$, then (2.2) becomes

$$L(u) = \sum_{j=1}^{s} P(C_j)\Lambda(u|C_j)$$

where

$$\Lambda(u|C_j) = \frac{1}{\mu}E\{YI[Y \leq F^{-1}(u)]|\mathbf{X} \in C_j\} = \frac{\mu(C_j)}{\mu}L(F_j(F^{-1}(u))|C_j), \tag{2.3}$$

and F_j, $\mu(C_j)$ and $L(\cdot|C_j)$ are the distribution function of Y, the mean of Y, and the Lorenz curve for sub-population C_j.

Note that $\sum P(C_j)\Lambda(1|C_j) = 1$, but $\Lambda(1|C_j) = \frac{\mu(C_j)}{\mu} \neq 1$ except when $\mu(C_j) = \mu$.

([5]) See Aaberge, Bjerve and Doksum (2005) who have used conditional Lorenz curves for deriving a regression framework for the Lorenz curve and the Gini coefficient.

The above decomposition of the Lorenz curve gives a method for identifying the contribution to overall inequality from each subgroup, where the subgroup contributions can be expressed as the product of three components; the proportion of the population that belongs to the subgroup, the ratio between the subgroup mean income, and the overall mean income and an interaction component that depends on income inequality within the subgroup as well as the relative location of the subgroup distribution.

Similar to (2.3) for the discrete case we get the following expression for the continuous case,

$$\Lambda(u|\mathbf{x}) = \frac{\mu(\mathbf{x})}{\mu} L(g(u)|\mathbf{x}) \tag{2.4}$$

where $g(u) = F(F^{-1}(u)|\mathbf{x})$ and $L(\cdot|\mathbf{x})$ is the (conditional) Lorenz curve for $F(y|\mathbf{x})$.

To summarize the information provided by the pseudo Lorenz curve $\Lambda(u|\mathbf{x})$ we may use the *pseudo-Gini coefficient*([6]) defined by

$$\Gamma(\mathbf{x}) = 2\int_0^1 [u - \Lambda(u|\mathbf{x})]du = \frac{1}{\mu} E\{Y[2F(Y) - 1]|\mathbf{x}\}, \tag{2.5}$$

or alternatively any member of the following family of pseudo inequality measures

$$\Psi_P(\mathbf{x}) = 1 + \int_0^1 P''(u)\Lambda(u|\mathbf{x})du = E\{Y[1 - P'(F(Y))]|\mathbf{x}\}, \tag{2.6}$$

where the weight-function P' is the derivative of a concave function P defined on the unit interval that satisfies the conditions $P(0) = 0$, $P(1) = 1$ and $P'(1) = 0$. Note that the unconditional counterpart of (2.6) is the family of rank-dependent measures of inequality introduced by Mehran (1976)([7]). By inserting for $P(u) = 2u - u^2$ in (2.5) we find that $\Psi_P(\mathbf{x}) = \Gamma(\mathbf{x})$. As for the pseudo-Lorenz curve we find the following convenient aggregation property for the pseudo inequality measures

$$\begin{aligned} J_P = E[\Psi_P(\mathbf{X})] &= EY[1 - P'(F(Y))] = 1 - \int_0^1 P'(u)dL(u) \\ &= 1 - \frac{1}{\mu}\int_0^1 P'(u)F^{-1}(u)du. \end{aligned} \tag{2.7}$$

([6]) Kakwani (1980) introduced a similar definition in cases where **x** is a vector of discrete variable. See also Mahalanobis (1960).
([7]) Mehran (1976) introduced the J_P-family by relying on descriptive arguments, whereas alternative normative motivations of the J_P-family and various subfamilies of the J_P-family have been provided by Donaldson and Weymark (1980, 1983), Weymark (1981), Yaari (1987,1988), Ben Porath and Gilboa (1994) and Aaberge (2001).

As demonstrated by Aaberge (2000) the Gini coefficient attaches an equal weight to a given transfer of income irrespective of where it takes place in the income distribution, as long as the income transfer occurs between individuals with the same difference in ranks. Thus, in general the Gini coefficient favors neither the lower nor the upper part of the Lorenz curve. To supplement the information provided by the Gini coefficient it might be relevant to use the Bonferroni coefficient([8]) defined by

$$B = \int_0^1 [1 - u^{-1} L(u)] du = 1 + \frac{1}{\mu} E[Y \log F(y)] \tag{2.8}$$

and the pseudo-Bonforroni coefficient defined by

$$B(\mathbf{x}) = \int_0^1 [1 - u^{-1} \Lambda(u|\mathbf{x})] du = 1 + \frac{1}{\mu} E\{[Y \log F(Y)]|\mathbf{x}\}. \tag{2.9}$$

Note that B and $B(\mathbf{x})$ corresponds to J_P and Ψ_P for $P(u) = u(1 - \log u)$. As demonstrated by Aaberge (2000) the Bonferroni coefficient B satisfies Mehran's principle of positional transfer sensitivity([9]) for any distribution function F and Kolm's principle of diminishing transfers for all F for which $\log F(x)$ is strictly concave. Thus, B is particular sensitive to transfers that occur in the lower part of the income distribution for logconcave distribution functions.

As suggested in Section 1 the main purpose of this paper is not to focus attention on the various components defined by the covariable vector **x** in cases where **x** is a vector of discrete variables, but to treat **x** as a vector of continuous variables and develop a framework that can be considered to provide similar information as the decomposition method in a situation with discrete variables. To this end we introduce the regression coefficients of the regression functions (2.1), (2.5) and (2.6) as quantities that provide information on the influence of covariates on overall inequality.

3. Measuring the effect of covariates on rank-dependent measures of inequality

By exploiting the parallel with the quantile regression approach, Aaberge, Bjerve and Doksum (2005) developed a regression framework for the conditional Lorenz curve, the conditional Gini coefficient and conditional rank-dependent measures of inequality, which can be used to examine the influence of covariates

([8]) For a discussion of the Bonferroni coefficient see D'Addario (1936), Nygård and Sandström (1981), Aaberge (1982, 2000) and Giorgi (1998). A poverty measure derived from the Bonferroni coefficient has been introduced by Giorgi (2001).
([9]) See also Nygård and Sandstrøm (1981) and Giorgi (1998).

x on income inequality in the conditional distribution ($F(y|\mathbf{x})$) of Y given given $\mathbf{X} = \mathbf{x}$. However, sine the overall Lorenz curve and the overall Gini coefficient will not be attained by averaging out the covariates in the conditional Lorenz curve and the conditional Gini coefficient, the effects of covariates on the conditional Lorenz curve and the conditional Gini coefficient do not immediately carry over to the overall Lorenz curve and the overall Gini coefficient. Thus, the (conditional) Lorenz and Gini regression coefficients are not the appropriate quantities when focus is turned to the effects of covariates on overall inequality. To this end it appears more relevant to consider the regression coefficients of the pseudo-Lorenz curve and the pseudo-Gini coefficient introduced in Section 2. The pseudo-Lorenz regression coefficient curves are defined by

$$\lambda_j(u;\mathbf{x}) = \frac{\partial \Lambda(u|\mathbf{x})}{\partial x_j}\,, \quad 0 \le u \le 1\,, \quad j = 1, 2, \dots, s\,, \tag{3.1}$$

and can be considered as measures of the relative importance of the covariate x_j on income inequality([10]). They show how much a small perturbation of x_j for $j = 1, 2, \dots, s$ changes the pseudo-Lorenz curves and allows the effects of the covariates to depend on whether the response is located in the lower, the central or the upper segment of the income distribution. Similarly as for the quantile regression coefficients curves it may be useful to summarize the pseudo-Lorenz regression coefficient curves across the covariates by

$$\lambda_j(u) = E\lambda_j(u;\mathbf{X})\,, \quad 0 \le u \le 1\,, \quad j = 1, 2, \dots, s\,. \tag{3.2}$$

Note that $\lambda_j(u)$ gives the average change of the pseudo-Lorenz curves due to small change in the j-th covariate when the remaining covariates are first kept fixed, then averaged out. We call $\lambda_j(\cdot)$ the j-th marginal pseudo-Lorenz curve.

To complete the summarization of the pseudo-Lorenz regression coefficients provided by $\lambda_j(u)$ a summary measure that captures the variation across quantiles will be introduced. To this end we may use the pseudo-Gini coefficient as a summary measure of the information content of the pseudo-Lorenz curve. The pseudo-Gini regression coefficients that correspond to (3.1) are defined by

$$\gamma_j(\mathbf{x}) = \frac{\partial \Gamma(\mathbf{x})}{\partial x_j} = -2\int_0^1 \lambda_j(u,\mathbf{x})du\,, \quad j = 1, 2, \dots, s\,. \tag{3.3}$$

Moreover, by summarizing over **x** we get

$$\gamma_j = E\gamma_j(\mathbf{X}) = -2\int_0^1 \lambda_j(u)du\,, \quad j = 1, 2, \dots, s\,. \tag{3.4}$$

([10]) A similar approach for quantile regression was introduced by Chaudhury *et al.* (1997).

The corresponding pseudo-Bonferroni summary measures are given by

$$b_j(\mathbf{x}) = \frac{\partial B(\mathbf{x})}{\partial x_j}\,, \quad b_j = Eb_j(\mathbf{X})\,. \tag{3.5}$$

Since alternative methods for summarizing the pseudo-Lorenz regression coefficients may be called for, we introduce the Ψ_P-regression coefficients derived from the pseudo-inequality measures defined by (2.5),

$$\xi_{jP}(\mathbf{x}) = \frac{\partial \Psi_P(\mathbf{x})}{\partial x_j} = \int_0^1 P''(u)\lambda_j(u,\mathbf{x})du\,, \quad j = 1, 2, \ldots, s\,, \tag{3.6}$$

where P'' is the second derivative of the weight-function P. By summarizing over $\mathbf{x}$ we get

$$\xi_{jP} = E\xi_{iP}(\mathbf{X}) = \int_0^1 P''(u)\lambda_j(u)du\,, \quad j = 1, 2, \ldots, s\,. \tag{3.7}$$

Note that $P(u) = 2u - u^2$ is the P-function that corresponds to the Gini coefficient, whilst $P(u) = u(1 - \log u)$ corresponds to the Bonferroni coefficient $(P''(u) = -1/u)$.

4. Estimation

We have considered a variety of maps $m : R^s \to R$ that measure inequality in income Y as a function of covariates $\mathbf{x} \in R^s$. These surfaces $m(\cdot)$, which are referred to as "curves" in the literature and this paper, can not be displayed effectively, nor estimated efficiently unless the sample sizes are enormous. For this reason we turn to summary measures: The average derivative nonparametric parameter is the expected value of the gradient vector

$$\nabla m(\mathbf{x}) = \left(\left[\frac{\partial}{\partial x_j} m(\mathbf{x})\right], j = 1, \ldots, s\right)^T\,. \tag{4.1}$$

In the case of single index models, $\nabla m(\mathbf{x})$ is proportional to the single index parameter vector.

Average Derivative Estimates (ADE's) have been proposed and analysed by Stoker (1986), Härdle and Stoker (1989), Härdle *et al.* (1993), Chaudhury *et al.* (1997), and Hristache *et al.* (2001), among others. Related work on projection pursuit regression appears in Friedman and Stuetzle (1981) and Hall

(1989). One basic idea is to estimate the gradient locally near a sample point $\mathbf{x}_i$ by using locally weighted least squares. That is, use $\hat{\nabla} m$, where

$$\begin{pmatrix} \hat{m}(\mathbf{X}_i) \\ \widehat{\nabla m}(\mathbf{X}_i) \end{pmatrix} = \arg \min_{a \in R} \min_{\beta \in R^s} \sum_{j=1}^{n} \{V_j - [a + \beta^T(\mathbf{X}_j - \mathbf{X}_i)]\}^2 K\left(\frac{|\mathbf{X}_j - \mathbf{X}_i|}{h^2}\right)$$
$$= \left\{\sum_{j=1}^{n} \begin{pmatrix} 1 \\ \mathbf{X}_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{X}_{ij} \end{pmatrix}^T K\left(\frac{|\mathbf{X}_{ij}|^2}{h^2}\right)\right\}^{-1} \sum_{j=1}^{n} V_j \begin{pmatrix} 1 \\ \mathbf{X}_{ij} \end{pmatrix} K\left(\frac{|\mathbf{X}_{ij}|^2}{h^2}\right). \tag{4.2}$$

Here h is a tuning parameter selected using the data, $\mathbf{X}_{ij} = \mathbf{X}_j - \mathbf{X}_i$, $|\cdot|$ is Euclidean distance, and the basic data $\{(\mathbf{X}_1, V_1), \ldots, (\mathbf{X}_n, V_n)\}$ is assumed to be i.i.d. The proceeding references give various modifications of this basic formula in order to deal with regions with sparse data.

To use these methods we need further specification. We have considered the following three m's:

$$\begin{aligned} m_1(u; \mathbf{x}) &= \Lambda(u|\mathbf{x}) = \mu^{-1} E\left[Y I[F(Y) \le u]|\mathbf{x}\right], \\ m_2(\mathbf{x}) &= \Gamma(\mathbf{x}) = 1 - 2\mu^{-1}\left[E(Y|\mathbf{x}) - E(Y F(Y)|\mathbf{x})\right], \\ m_3(\mathbf{x}) &= B(\mathbf{x}) = 1 + \mu^{-1} E\left[Y \log F(Y)|\mathbf{x}\right]. \end{aligned} \tag{4.3}$$

Thus, we need ADE's for the four cases where $V = YI[F(Y) \le u]$, $V = Y$, $V = YF(Y)$ and $V = Y \log F(Y)$.

Because F is unknown, we need to replace $F(Y_i)$ by its empirical version $\hat{F}(Y_i) = i/n$, where the incomes $\{Y_i\}$ have been arranged in increasing order and $\mathbf{X}_i$ now denotes the covariate vector that belongs with the i-th ordered Y. For ease of interpretation and display the ADE algorithms require that each X_{ij} in the sample have the sample mean $\overline{X}_j$ subtracted and be divided by the sample standard deviation s_j, $j = 1, \ldots, s$. Our curves require an estimate of $\mu = E(Y)$, which we take as $\hat{\mu} = \overline{Y}$.

We label the outputs from the ADE algorithms as $\widehat{\nabla m}_{kj}(\mathbf{X}_i)$, $k = 1, 2, 3$, $j = 1, \ldots, s$, $i = 1, \ldots, n$. Then our estimates are

$$\begin{aligned} \hat{\lambda}_j(u) &= n^{-1} \sum_{i=1}^{n} \widehat{\nabla m}_{1j}(u; \mathbf{X}_i) \text{ (Lorenz curve in direction } X_j) \\ \hat{\gamma}_j &= n^{-1} \sum_{i=1}^{n} \widehat{\nabla m}_{2j}(\mathbf{X}_i) \quad \text{(Gini coefficient in direction } X_j) \\ \hat{b}_j &= n^{-1} \sum_{i=1}^{n} \widehat{\nabla m}_{3j}(\mathbf{X}_i) \quad \text{(Bonferroni coefficient in direction } X_j) \end{aligned} \tag{4.4}$$

When there is only one covariate X, estimation is more straightforward. In the case of $\Lambda(u|\mathbf{x})$ we can apply any nonparametric regression estimator to the data $(X_1, V_1(u), \ldots, X_n, V_n(u))$ where

$$V_i(u) = I(i \leq [un])Y_i \tag{4.5}$$

and $[\cdot]$ is the greatest integer function. One simple such estimator would be

$$\hat{\Lambda}(u|x) = \frac{\sum_{i=1}^{n} V_i(u) K_h(X_i - x)}{\overline{Y} \sum_{i=1}^{n} K_h(X_i - x)} \tag{4.6}$$

where $K_h(u) = h^{-1}K(u/h)$, $K(u)$ is a kernel on R with $\int K(u)du = 1$ and $h > 0$ is a tuning parameter.

The Gini regression index can be estimated as

$$\hat{\Gamma}(x) = 1 - 2(\overline{Y})^{-1}\hat{\mu}_G(x) \tag{4.7}$$

where

$$\hat{\mu}_G(x) = \frac{\sum_{i=1}^{n} \left(1 - \frac{i}{n+1}\right) Y_i K_h(X_i - x)}{\sum_{i=1}^{n} K_h(X_i - x)} \tag{4.8}$$

Here the $\{Y_i\}$ are in increasing order and X_i is the covariate value for the case with ordered response Y_i. Similarly, the Bonferroni regression index can be estimated as

$$\hat{B}(x) = 1 + (\overline{Y})^{-1}\hat{\mu}_B(x) \tag{4.9}$$

where

$$\hat{\mu}_B(x) = \frac{\sum \log\left(\frac{i}{n+1}\right) Y_i K_L(X_i - x)}{\sum K_L(X_i - x)} \tag{4.10}$$

ACKNOWLEDGMENTS

We would like to thank Giovanni Maria Giorgi for helpful comments. Steinar Bjerve gratefully acknowledges the support of The Wessmann Society during the course of this work.
Kjell Doksum's work was supported in part by NSF grants DMS-9971301 and DMS-0505651.

REFERENCES

AABERGE, R. (1982) Om måling av ulikskap, *Rapporter 82/9*, Statistisk sentralbyrå (In Norwegian).

AABERGE, R. (2000) Characterizations of Lorenz Curves and Income Distributions, *Social Choice and Welfare*, 17, 639–653.

AABERGE, R. (2001) Axiomatic Characterization of the Gini Coefficient and Lorenz Curve Orderings, *Journal of Economic Theory*, 101, 115–132.

AABERGE, R, BJERVE S., and DOKSUM, K. (2005) *Lorenz and Gini Regression*, Mimeo, Research Department, Statistics Norway.

BEN PORATH, E. and GILBOA, I. (1994) Linear Measures, the Gini Index, and the Income-Equality Trade-off, *Journal of Economic Theory*, 64, 443–467.

BHATTACHARYA, N. and MAHALANOBIS, B. (1967) Regional Disparities in Household Consumption in India, *JASA*, 62, 143–161.

BICKEL, P. J. and DOKSUM, K. D. (2001) *Mathematical Statistics. Basic Ideas and Selected Topics*, 2nd Ed., Vol. I. New Jersey: Prentice Hall.

CHAUDHURI, P., DOKSUM, K. A., and SAMAROV, A. (1997) On Average Derivative Quantile Regression, *Journal of the American Statistical Association*, 74, 829–836.

D'ADDARIO, R. (1936) *Curve di concentrazione, elasticitá, flessibilitá, densitá media e densitá marginale dei redditi*, Cressati, Bari.

DAGUM, C. (1997) A new Approach to the Decomposition of the Gini Income Inequality Ration, *Empirical Economics*, 22, 515–531.

DONALDSON, D. and WEYMARK, J. A. (1980) A Single Parameter Generalization of the Gini Indices of Inequality, *Journal of Economic Theory*, 22, 67–86.

DONALDSON, D. and WEYMARK, J. A. (1983) Ethically Flexible Indices for Income Distribution in the Continuum, *Journal of Economic Theory*, 29, 353–358.

FRIEDMAN, F. H. and STUETZLE, W. (1981) Projection Pursuit Regression, *Journal of the American Statistical Association*, 76, 817-823.

GIORGI, G. M. (1990) Bibliographic portrait of the Gini concentration ratio, *Metron*, 48, 183–221.

GIORGI, G. M. (1998) Concentration index, Bonferroni, In: *Encyclopedia of Statistical Sciences*, Kotz S., Read C.B., and Banks D.L. (eds.) Update, 2, 141–146.

GIORGI, G. M.(2001) A proposal of poverty measures based on the Bonferroni inequality index, *Metron*, 59 (3-4), 3–16.

HALL, P. (1989) On Projection Pursuit Regression, *Annals of Statistics*, 17, 573–588.

HRISTACHE, M., JUDITSKY, A., and SPOKOINY V. (2001) Direct Estimation of the Index Coefficient in a Single-Index Model, *Annals of Statistics*, 17, 1009–1052.

HÄRDLE, W. and STOKER, T. (1989) Investigating Smooth Multiple Regression by the Method of Average Derivatives, *Journal of the American Statistical Association*, 84, 986–995.

HÄRDLE, W. P. HALL and ICHIMURA, H. (1993) Optimal Smoothing in Single-Index Models, *Annals of Statistics*, 21, 157–178.

KAKWANI, N. C. (1980) *Income Inequality and Poverty: Methods of Estimation and Policy Applications*, Oxford University Press.

LAMBERT, P. and ARONSON, R. (1993) Inequality Decomposition Analysis and the Gini Coefficient Revisited, *The Economic Journal*, 103, 1221–1227.

LERMAN, R. I. and YITZHAKI, S. (1985) Income Inequality Effects by Income Source: A New Approach and Applications to the United States, *Rewiew of Economics and Statistics*, 67, 151–156.

MAHALANOBIS, P. C. (1960) A Method of Fractile Graphical Analysis, *Econometrica*, 28, 325–351.

MEHRAN, F. (1976) Linear measures of inequality, *Econometrica*, 44, 805–809.

NYGÅRD, F. and SANDSTRØM, A. (1981) *Measauring Income Inequality*, Almqvist and Wiksell, Stockholm.

PIESCH, W. (1975) *Statistische Konzentrationsmasse*, Mohr (Siebeck), Tübingen.

SILBER, J. (1989) Factor Components, Population Subgroups, and Computation of the Gini Index of Inequality, *Review of Economics and Statistics*, 71, 107–115.

SHORROCKS, A. F. (1984) Inequality Decomposition by Population Subgroups, *Econometrica*, 52, 1359–1385.

STOKER, T. (1986) Consistent Estimation of Scaled Coefficients, *Econometrica*, 54, 1461–1481.

WEYMARK, J. (1981) Generalized Gini Inequality Indices, *Mathematical Social Science*, 1, 409–430.

YITZHAKI, S. (1994) Economic Distance and Overlapping of Distribution, *Journal of Economics*, 61, 147–159.

YITZHAKI, S. and LERMAN, R.I. (1991) Income Stratification and Income Inequality, *Review of Income and Wealth*, 37, 313-329.

YAARI, M. E. (1987) The Dual Theory of Choice under Risk, *Econometrica*, 55, 95–115.

YAARI, M. E. (1988) A Controversial Proposal Concerning Inequality Measurement, *Journal of Economic Theory*, 44, 381–397.

ROLF AABERGE
Research Department
Statistics Norway
P.O. Box 8131 Dep.
N-0033 Oslo (Norway)
rolf.aaberge@ssb.no

STEINAR BJERVE
Department of Mathematics
University of Oslo
P.O.Box 1053
N-0316 Oslo (Norway)
steinar@math.uio.no

KJELL DOKSUM
Department of Statistics
University of Wisconsin-Madison
1180 Observatory Drive, Madison
Wisconsin 53706 (U.S.A.)
doksum@stat.wisc.edu