*Stochastic Orders and Decision under Risk*
IMS Lecture Notes – Monograph Series (1991).

# CONCENTRATION INDICES AND CONCENTRATION CURVES

BY SHLOMO YITZHAKI AND INGRAM OLKIN *

*Hebrew University and Stanford University*

For analyzing the relationship between two random variables an approach is introduced which is called the Gini method. The approach is based on the Gini mean difference, the Gini covariance, and the Gini correlation. The method is then extended to include concentration curves. For two given random variables a condition in terms of their concentration curves (with respect to themselves) is derived which is necessary and sufficient for second degree stochastic dominance between the variables.

**1. Introduction.** The last two decades have witnessed a revival of interest in the Gini coefficient and the Lorenz curve. Authors in different fields independently discovered the usefulness of the Lorenz curve, the concentration curve and the Gini coefficient. Atkinson (1970) showed that the rules for ordering risky prospects can be presented in a simple way by Lorenz curves (see Hadar and Russel (1969), Hanoch and Levy (1969), Rothschild and Stiglitz (1970)). These rules are referred to in the finance literature as *stochastic dominance rules.*

Atkinson's paper was followed by papers utilizing the Lorenz curve (and the Gini coefficient which is a summary statistic based on it) both in the field of finance and in the more traditional field of income distribution. In finance, rules were developed for ordering risky prospects utilizing the Gini index (Yitzhaki (1982)) and also for the evaluation of risky assets (Shalit and Yitzhaki (1984)). These rules are on the one hand similar to the classical rules using the variance (mean–variance rules, and capital asset pricing model, Markowitz (1952) and Sharpe (1964)), and on the other hand they rank prospects in accordance with the maximization of expected utility even when prospects are not normally distributed.

---

* Supported in part by National Science Foundation Grant DMS 90-02502.

AMS 1980 Subject Classification: Primary 90A19; Secondary 62H20.

Key words and phrases: Correlation, Gini mean difference, majorization, stochastic dominance, Gini coefficient, Lorenz curves.

Although the statistical properties of these indices are not fully developed, this theoretical literature is beginning to be followed by an empirical one (Bey and Howe (1984), Buccola and Subai (1984), Okunev (1988)). In the field of income distribution there is a literature on the decomposition of the Gini coefficient by income sources and sub-populations (Rao (1969), Fields (1979), Pyatt, Chen and Fei (1980), Das and Parikh (1982), Sandström (1982), Shorrocks (1983a), Zagier (1983), Lerman and Yitzhaki (1985)), the extension of the Gini coefficient into a family of indices of inequality, depending on one parameter (Donaldson and Weymark (1980, 1983), Yitzhaki (1983)), which enables the investigator to stress parts of the distribution and the development of the concept of generalized Lorenz curve (Shorrocks (1983b), Atkinson and Bourguignon (1989)) for welfare comparisons. An important branch of this literature deals with poverty (Sen (1976)). This literature extends the methods developed in income distributions to truncated distributions.

In a different line of research Kakwani (1977, 1980), following an unnoticed paper of Mahalanobis (1960), showed that the Lorenz curve and the concentration curve enable the investigator to nonparametrically estimate the income elasticity of a consumption good. This method can be used in the study of elasticity of consumption with respect to income and progression of taxes (Jakobsson (1976), Suits (1977)).

One of the obstacles in the use of these tools is the lack of an exposition of a systematic development of the theory. Actually, many of these tools previously used were developed independently in different fields [see, for example, Simpson (1949), David (1968), Chandra and Singpurwalla (1981)].

One consequence of the independent development in different fields is the lack of a unified terminology, whereby different words have been used for similar concepts. Therefore it behooves us to set forth some of the terminology used in this paper. Our purpose is to provide a systematic development of concentration curves and the Gini method. The main property of concentration curves is that they enable us to form necessary and sufficient conditions for various forms of the second-degree stochastic dominance criteria. However, any empirical application of these criteria will be cumbersome, since comparisons of many curves are required. Hence a screening device or an index is needed in order to reduce the set of possible candidates for dominances. The screening device is supplied by the Gini method, which forms necessary conditions for dominance. Therefore it would be reasonable to first analyze the data using the Gini method. If the necessary conditions for stochastic dominance are fulfilled, then one should carry out a subsequent analysis based on concentration curves.

This paper consists of two main thrusts. The first (Sections 2–3) is a discussion of the Gini mean difference. This development provides some insight into how alternative extensions may be formulated. The second thrust (Section 5) extends the ideas of the Gini Method to concentration curves. This latter section contains a number of new concepts heretofore not discussed. It would be beyond the scope of a single paper to give a complete set of references. Instead we have selected a number of references from which a more complete set can be obtained.

**2. Preliminaries and Terminology.** There are two standard approaches for analyzing the relationship between two variables, $X$ and $Y$, both based on the properties of the covariance. In one the variates themselves are used, whereas in the other cumulative distributions are used. If we use the variates, then the key parameter is $\mathrm{Cov}(X,Y)$. This leads to a measure of variability of one variable as $\mathrm{Cov}(X,X)$, and a Pearson's correlation coefficient emerges as a standardized covariance. Hereafter, this method is referred to as the *variate method*. The other approach is to consider the covariance between the cumulative distributions, $\mathrm{Cov}(F(X),G(Y))$, where $F$ and $G$ are the marginal distributions of $X$ and $Y$, respectively. This leads to Spearman's correlation coefficient as the standardized covariance.

Some statisticians (e.g., Daniels (1944), Stuart (1954), Kendall (1955), Barnett, Green and Robinson (1976)) use a third method which is a mixture of those mentioned above, namely, $\mathrm{Cov}(X,G(Y))$, the covariance between $X$ and the cumulative distribution of $Y$. As we show later, this parameter can be studied in several ways, and therefore the terminology used with regard to statistics based on it is not uniform and may at times be confusing. Throughout the paper we confine ourselves to the case of absolutely continuous distributions. This narrows severely the applicability of the results, but renders the mathematics much simpler.

Analogous to the variate method, $\mathrm{Cov}(X,F(X))$ is a measure of variability, and

$$\frac{\sqrt{|\mathrm{Cov}(X,G(Y))\,\mathrm{Cov}(Y,F(X))|}}{\sqrt{\mathrm{Cov}(X,F(X))\,\mathrm{Cov}(Y,G(Y))}}$$

may be defined as an index of correlation that ranges between 0 and 1.

An alternative measure that ranges between $-1$ and $+1$ is

$$\frac{1}{2}\frac{\mathrm{Cov}(X,G(Y))}{\mathrm{Cov}(X,F(X))}+\frac{\mathrm{Cov}(Y,F(X))}{\mathrm{Cov}(Y,G(Y))}.$$

Both definitions are symmetric functions of $X$ and $Y$. However, an asymmetric version provides a more suitable interpretation for our present purposes:

$$\rho_{Y \circ X} = \frac{\operatorname{Cov}(Y, F(X))}{\operatorname{Cov}(Y, G(Y))}. \tag{2.1}$$

Such an asymmetry is intrinsic in regression analysis but not in correlation analysis. Because $\operatorname{Cov}(X, F(X))$ is proportional to Gini's Mean Difference (shown later). We will refer to $\operatorname{Cov}(Y, F(X))$ as the Gini covariance and $\rho_{Y \circ X}$ as the Gini correlation. The composite of the Gini Index, the Gini covariance and the Gini correlation is referred to as the *Gini method.* The term *Gini method* is intended to signify that the method is based on Gini's mean difference, and does not reflect an approach that Gini might have used.

The Gini method has some of features of other methods. When $X$ and $Y$ have a bivariate normal distribution,

$$\operatorname{Cov}(X, G(Y)) = \sigma_{XY}/2\sigma_Y\sqrt{\pi}, \tag{2.2}$$

so that

$$\rho_{X \circ Y} = \frac{\operatorname{Cov}(X, G(Y))}{\operatorname{Cov}(X, F(X))} = \frac{\operatorname{Cov}(Y, F(X))}{\operatorname{Cov}(Y, G(Y))} = \rho_{Y \circ X} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \rho, \tag{2.3}$$

which is the usual product-moment correlation coefficient. (See Schechtman and Yitzhaki (1987).) On the other hand, the Gini correlation coefficient, $\rho_{Y \circ X}$, is not affected by a monotonic transformation of $X$, a property shared with Spearman's correlation coefficient. Also, the regression coefficient of $Y$ on $X$ is

$$\beta_{Y.X} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\operatorname{Cov}(Y, F(X))}{\operatorname{Cov}(X, F(X))},$$

which, similar to the ordinary least squares regression coefficient, is a cardinal measure. Estimators of the parameters of the Gini method are asymptotically distribution free, which may be considered to be between parametric and nonparametric estimation obtained with other approaches.

In general, the Gini covariance and correlation are not symmetric in $X$ and $Y$. Moreover, $\operatorname{Cov}(X, G(Y))$ and $\operatorname{Cov}(Y, F(X))$ may have different signs. This property may be viewed as a deficiency of the method. On the other hand, there are important instances of asymmetric concepts, as for example regression analysis and the concept of elasticity in economics. However, the Gini correlation is symmetric for exchangeable variables, and hence can be used for constructing tests for asymmetry (Boos (1982)).

As we show later, all parameters of the Gini method can be presented as parameters describing certain properties of concentration curves. Alternatively, one can view concentration curves as an extension of the parameters of the Gini method, which enables the investigator to see how the coefficients are composed. Hence, whenever a coefficient is sufficient for the analysis, the use of the Gini method is appropriate. On the other hand, there are some fields such as stochastic dominance where coefficients do not provide sufficient tools for the analysis. In these cases one has to use the concentration curves.

**3. Alternative Representations of Gini's Mean Difference.** Gini's mean difference has been known for more than a century (see e.g., David, 1981, p. 192), and has several representations. Therefore it is not surprising that it was reinvented by several authors, often using different terminology (see Simpson (1949), David (1968)).

In this section we present some definitions of the Gini index which serve to motivate some of the applications. Although most definitions are useful in one dimension, because one of our concerns is to define a multivariate version, [see Taguchi (1972a, 1972b, 1973) for multivariate definitions of the Gini coefficient and Lorenz curve and Atkinson and Bourguignon (1982) for multivariate Lorenz curves], some definitions will be more adaptable for such extensions.

First, there is the definition of Gini's index $\mathcal{G}$ as the expected difference (taken regardless of sign) of each possible pair of variate values. Formally, if $X_1$ and $X_2$ are independent identically distributed random variables, then

$$\mathcal{G} = \mathcal{E}|X_1 - X_2|/2. \tag{3.1}$$

Although this is a simple expression, it is not always easy to handle and is used infrequently. However, this definition has an intuitive interpretation as a concept of variability in that it is one half the expected difference between two randomly drawn realizations of the variable.

Another definition of the Gini index $\mathcal{G}$ is

$$\begin{aligned}\mathcal{G} &= \mu - \mathcal{E}\{\min(X_1, X_2)\} = \mathcal{E}\{\max(X_1, X_2)\} - \mu \\ &= \frac{1}{2}\mathcal{E}\{\max(X_1, X_2) - \min(X_1, X_2)\},\end{aligned} \tag{3.2}$$

where $\mu = \mathcal{E}X_1 = \mathcal{E}(X_1 + X_2)/2$. These versions of the Gini index are useful in applications involving extreme value theory for which the extrema play a special role and may be more readily computed. (See Gumbel (1958).)

A third definition of the Gini index which has considerable promise is

$$\mathcal{G} = 2\,\mathrm{Cov}(X, F(X)) = -2\mathrm{Cov}(X, \bar{F}(X)), \tag{3.4}$$

where $\bar{F}(x) = P\{X > x\} = 1 - F(x)$. This shows more clearly how the Gini index is a compromise between the variate and rank covariances.

By using (3.4) the Gini index can be calculated from a single regression program, which incidentally yields a geometric interpretation of the Gini index (Lerman and Yitzhaki (1984)). The equivalence of (3.1) and (3.4) was shown by Stuart (1954) and by Lerman and Yitzhaki (1984).

An alternative definition used in the literature is

$$\mathcal{G} = \int_{-\infty}^{\infty} F(x)\bar{F}(x)\,dx, \tag{3.7}$$

whose extension to the multivariate case is clear, where now $\bar{F}(x_1,\ldots,x_\nu) = P\{X_1 > x_1,\ldots,X_\nu > x_\nu\}$.

The Gini coefficient is the most known member of the parameters of the Gini method. It can be presented as:

$$\Gamma = 2\mathrm{Cov}\big(X, F(X)\big)/\mu_X.$$

The Lorenz curve was originally defined by Lorenz (1905) as "plot along one axis cumulated per cent of the population from the poorest to the richest, and along the other axis the percent of the total wealth held by these percents of the population." This curve enables us to obtain more information on the variability of the variable, which may be important whenever a summary statistic is not sufficient. The extension of the Gini coefficient to Gini's mean difference can be paralleled by extending the Lorenz curve to the *Absolute Concentration Curve* and to the absolute concentration curve of one variable with respect to another variable (parallel to the Gini covariance). The term *concentration curve* is borrowed from the income distribution literature, where these curves were developed. We discuss this formulation when we deal with concentration curves.

When the underlying distribution is normal $\mathcal{G} = \sigma/\sqrt{\pi}$, so that one can use the Gini index to estimate the variance. The advantage of this procedure is that the Gini index is less sensitive to outliers than is the variance; the disadvantage is the loss of efficiency. A discussion of the efficiency of the Gini index is provided, for example, by Nair (1936), Lomnicki (1952).

As shown by Sievers (1983a) several alternative estimates to regression coefficients actually minimize the Gini index of the error term in the regression (see also Jurečková (1969, 1971), Jaeckel (1972), McKean and Hettmansperger (1978)). Sievers (1983b) who discovered the Gini index in regressions developed a concept which can be termed the Gini multiple correlation coefficient. It is an extension of the Gini correlation, and it plays a role similar to that of the multiple correlation coefficient.

**4. Absolute Concentration Curves.** We here adopt the following standard notation. A bivariate density function of $(X, Y)$ is denoted by $h(\cdot,\cdot)$; its distribution function is $H(\cdot,\cdot)$. The marginal densities and distribution functions are denoted by $f(\cdot)$, $g(\cdot)$, $F(\cdot)$, and $G(\cdot)$, respectively. The expected values are $\mu_X$ and $\mu_Y$. The conditional density function of $Y$ given $X = x$ is $h_{Y|x}$ and the conditional expectation $E(Y \mid X = x)$ is $m(x)$.

Besides the regularity conditions mentioned in Section 2 from now on we will assume the differentiability of all the functions encountered. Furthermore, the distribution functions will be strictly increasing.

*4.1. Definitions of Absolute and Relative Concentration Curves.* The *absolute concentration curve* (ACC) of $Y$ with respect to $X$ denoted $A_{Y\circ X}(\theta)$ is defined by the relationship

$$A_{Y\circ X}(\theta) = \int_{-\infty}^{F^{-1}(\theta)} m(t)dF(t) = E(YI_{X\leq F^{-1}(\theta)}) \quad 0 \leq \theta \leq 1$$

where $I$ denotes the indicator function. The relative concentration curve (RCC) of $Y$ with respect to $X$ denoted by $R_{Y\circ X}$ is defined by

$$R_{Y\circ X}(\theta) = \frac{1}{\mu_Y} A_{Y\circ X}(\theta), \quad \mu_Y > 0.$$

The Lorenz curve is the special case of the relative concentration curve given by $R_{Y\circ Y}$, and we call $A_{Y\circ Y}$ the *absolute Lorenz curve* (called the Generalized Lorenz Curve (Shorrocks, 1983b)). For a general discussion of the properties of concentration curves and the connection to inequality measures, see Kakwani (1980) and Nygård and Sandström (1981).

*4.2. Properties of the Absolute and Relative Concentration Curves.* For simplicity of exposition, we write ACC instead of $A_{Y\circ X}(F)$ for the absolute concentration curve and RCC instead of $R_{Y\circ X}$ for the relative concentration curve. Whenever RCC is referred to, it is assumed that $\mu_Y > 0$.

FACT 4.1. The ACC passes through the points $(0,0)$ and $(1,\mu_Y)$; the RCC passes through the points $(0,0)$ and $(1,1)$.

FACT 4.2. The derivative of the ACC with respect to $\theta$ is $m(F^{-1}(\theta))$. As a consequence $A_{Y\circ X}$ is increasing if and only $m$ is positive.

FACT 4.3. The ACC and RCC are concave if and only if $m$ is nondecreasing.

FACT 4.4. If $\mathcal{E}(Y|X = x)$ is constant then the ACC is a straight line.

For any concentration curve, connect the points $(0,0)$ and $(\mu_Y, 1)$ by a straight line; we call this line the *line of independence* (LOI). If $X$ and $Y$ are independent then $A_{Y\circ X}$ is a straight line and coincides with the LOI. For the

RCC, the LOI is the 45° line connecting $(0,0)$ to $(1,1)$. In the case of the Lorenz curve the LOI is known as the *egalitarian line.*

**5. Majorization (Stochastic Dominance).** The rules of stochastic dominance were developed in the literature of finance by Roy (1953), Hadar and Russel (1969), Hanoch and Levy (1969), Rothschild and Stiglitz (1970). For a review of the stochastic dominance literature see Kroll and Levy (1980). The connection between those rules and the Lorenz curve was discovered by Atkinson (1970). Similar rules were developed independently in statistics by Marshall and Olkin (1979). Our interest is focused on the *second degree stochastic dominance criterion* (SSD): $X$ dominates $Y$ according to second order stochastic dominance criterion, denoted $X \underset{SSD}{\succ} Y$ if

$$\mathcal{E}\{U(X)\} \geq \mathcal{E}\{U(Y)\}$$

for all nondecreasing continuous concave functions $U(\cdot)$ such that the expectations exist. In the terminology of Marshall and Olkin (1979) this criterion is called "$X$ is weakly majorized by $Y$".

There is a connection between second order stochastic dominance criteria and ACC.

THEOREM 5.1. *Let $E(X)$, $E(Y)$ exist finite. $A_{X \circ X}(\theta) \geq A_{Y \circ Y}(\theta)$ $\forall$ $\theta$ if and only if $X \underset{SSD}{\succ} Y$.*

PROOF. A necessary and sufficient condition for $X \underset{SSD}{\succ} Y$ is

$$\int_{-\infty}^{z} F(t)dt \leq \int_{-\infty}^{z} G(t)dt \quad \forall\, z \in \mathbb{R}$$

(see e.g. Marshall and Olkin (1979) and Hanoch and Levy (1969). Let's prove the "only if" part first. If $A_{X \circ X}(\theta) \geq A_{Y \circ Y}(\theta)$ $\forall$ $\theta$, then

$$D \equiv \int_{-\infty}^{F^{-1}(\theta)} t dF(t) - \int_{-\infty}^{G^{-1}(\theta)} t dG(t) \geq 0.$$

The finiteness of $E(X)$ and $E(Y)$ implies

$$\lim_{t \to -\infty} tF(t) = \lim_{t \to -\infty} tG(t) = 0.$$

Therefore

$$D = \theta(F^{-1}(\theta) - G^{-1}(\theta)) - \int_{-\infty}^{F^{-1}(\theta)} F(t)dt + \int_{-\infty}^{G^{-1}(\theta)} G(t)dt \geq 0.$$

We need to consider three cases: (i) $F^{-1}(\theta) > G^{-1}(\theta)$, (ii) $F^{-1}(\theta) = G^{-1}(\theta)$, (iii) $F^{-1}(\theta) < G^{-1}(\theta)$.

If (i) holds, then

$$(F^{-1}(\theta) - G^{-1}(\theta))\theta - \int_{G^{-1}(\theta)}^{F^{-1}(\theta)} G(t)dt \geq \int_{-\infty}^{F^{-1}(\theta)} [F(t) - G(t)]dt.$$

Since $G$ is increasing, the left hand side is nonpositive.

The proof for case (ii) is trivial.

Finally, suppose that (iii) holds. Then

$$D = (F^{-1}(\theta) - G^{-1}(\theta))\theta - \int_{-\infty}^{F^{-1}(\theta)} [F(t) - G(t)]dt + \int_{F^{-1}(\theta)}^{G^{-1}(\theta)} G(t)dt \geq 0.$$

But

$$\begin{aligned}(F^{-1}(\theta) - G^{-1}(\theta))\theta &+ \int_{F^{-1}(\theta)}^{G^{-1}(\theta)} G(t)dt \\ &\leq (F^{-1}(\theta) - G^{-1}(\theta))\theta + (G^{-1}(\theta) - F^{-1}(\theta))\theta = 0.\end{aligned}$$

To prove the "if" part, suppose that for all $z$

$$\int_{-\infty}^{z} F(t)dt \leq \int_{-\infty}^{z} G(t)dt.$$

Integration by parts yields

$$zF(z) - \int_{-\infty}^{z} tdF(t) \leq zG(z) - \int_{-\infty}^{z} tdG(t). \tag{5.1}$$

Again we consider three cases: (i) $F^{-1}(\theta) > G^{-1}(\theta)$, (ii) $F^{-1}(\theta) = G^{-1}(\theta)$, (iii) $F^{-1}(\theta) < G^{-1}(\theta)$.

For case (iii) let $z = G^{-1}(\theta)$ in (5.1) to yield

$$\int_{-\infty}^{G^{-1}(\theta)} tdF(t) - \int_{-\infty}^{G^{-1}(\theta)} tdG(t) \geq G^{-1}(\theta)F(G^{-1}(\theta)) - G^{-1}(\theta)\theta,$$

or equivalently, with the use of integration by parts,

$$\int_{-\infty}^{F^{-1}(\theta)} tdF(t) - \int_{-\infty}^{G^{-1}(\theta)} tdG(t) \geq G^{-1}(\theta)[F(G^{-1}(\theta)) - \theta] - \int_{F^{-1}(\theta)}^{G^{-1}(\theta)} tdF(t)$$

$$= G^{-1}(\theta)[F(G^{-1}(\theta)) - \theta] - G^{-1}(\theta)F(G^{-1}(\theta)) + F^{-1}(\theta)\theta + \int_{F^{-1}(\theta)}^{G^{-1}(\theta)} F(t)dt$$

$$= \theta[F^{-1}(\theta) - G^{-1}(\theta)] + \int_{F^{-1}(\theta)}^{G^{-1}(\theta)} F(t)dt$$

$$= \int_{F^{-1}(\theta)}^{G^{-1}(\theta)} [F(t) - F(F^{-1}(\theta))]dt \geq 0.$$

Case (i) is parallel to case (iii) with $z = F^{-1}(\theta)$ in (5.1). Case (ii) is immediate. ∎

Necessary conditions on stochastic dominance can be derived using the Gini method (Yitzhaki (1982, 1983)).

## REFERENCES


ATKINSON, A. B. (1970). On the measurement of inequality. *J. Econom. Theory* **2**, 244–263.

ATKINSON, A. B. and BOURGUIGNON, F. (1982). The comparison of multi-dimensioned distributions of economic status. *Rev. Econom. Stud.* **49**, 183–201.

ATKINSON, A. B. and BOURGUIGNON, F. (1989). The design of direct taxation and family benefits. *J. Publ. Econom.* **41**, 3–29.

BARNETT, V., GREEN, P. J. and ROBINSON, A. (1976). Concomitants and correlation estimates. *Biometrika* **63**, 323–328.

BEY, R. P. and HOWE, K. M. (1984). Gini's mean difference and portfolio selection: An empirical evaluation. *J. Finan. and Quant. Anal.* **19**, 329–338.

BOOS, D. (1982). A test for asymmetry associated with the Hodges- Lehman estimator. *J. Amer. Statist. Assoc.* **77**, 647–651.

BUCCOLA, S. T. and SUBAI, A. (1984) Mean Gini analysis, stochastic efficiency and weak risk aversion. *Austral. J. Agri. Econom.* **28**, 77–86.

CHANDRA, M. and SINGPURWALLA, N. D. (1981). Relationships between some notions which are common to reliability theory and economics. *Math. Oper. Res.* **6**, 113–121.

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika* **33**, 129–135.

DAS, T. and PARIKH, A. (1982). Decomposition of Inequality Measures and a Comparative Analysis, *Empir. Econom.* **7**, 23–48.

DAVID, H. A. (1968). Gini's mean difference rediscovered. *Biometrika* **55**, 573–575.

DAVID, H. A. (1981). *Order Statistics* (2nd edition), New York: John Wiley & Sons.

DONALDSON, D. and WEYMARK, J. A. (1980). A single-parameter generalization of the Gini indices of inequality. *J. Econom. Theory* **22**, 67–86.

DONALDSON, D. and WEYMARK, J. A. (1983). Ethically flexible Gini indices for income distributions in the continuum. *J. Econom. Theory* **29**, 353–358.

FIELDS, G. S. (1979). Income Inequality in Urban Columbia: A Decomposition Analysis, *Rev. Income and Wealth*, 327–41.

GUMBEL, E. J. (1958). *Statistics of Extremes*, New York: Columbia University Press.

HADAR, J. and RUSSELL, W. R. (1969). Rules for ordering uncertain prospects. *Amer. Econom. Rev.* **59**, 25–34.

HANOCH, G. and LEVY, H. (1969). The efficiency analysis of choices involving risk. *Rev. Econom. Stud.* **36**, 335–346.

JAECKEL, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43**, 1449–1458.

JAKOBSSON, U. (1976). On the measurement of degree of progression. *J. Publ. Econom.* **5**, 161–168.

JUREČKOVÁ, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* **40**, 1889–1900.

JUREČKOVÁ, J. (1971). Nonparametric estimates of regression coefficients. *Ann. Math. Statist.* **42**, 1328–1338.

KAKWANI, N. C. (1977). Applications of Lorenz curves in economic analysis. *Econometrica* **45**, 719–727.

KAKWANI, N. C. (1980). *Income Inequality and Poverty*. Oxford: Oxford University Press.

KENDALL, M. G. (1955). *Rank Correlation Methods* (Second Edition). New York: Hafner Publishing Company.

KROLL, Y. and LEVY, H. (1980). Stochastic dominance: A review and some new evidence. Pages 163–227 in *Research in Finance*, Vol. 2, (edited by H. Levy), Greenwich: JAI Press.

LERMAN, R. I. and YITZHAKI, S. (1984). A note on the calculation and interpretation of the Gini index. *Econom. Lett.* **15**, 363–368.

LERMAN R. I. and YITZHAKI, S. (1985). Income inequality effects by income source: A new approach and applications to the U.S. *Rev. Econom. Statist.* **LXVII**, 151–156.

LOMNICKI, Z. A. (1952). The standard error of Gini's mean difference. *Ann. Math. Statist.* **23**, 635–636.

LORENZ, M. O. (1905). Methods for measuring concentration of wealth. *J. Amer. Statist. Assoc.*, New Series No. 70, 1905.

MAHALANOBIS, P. C. (1960). A method of fractile graphical analysis. *Econometrica* **28**, 325–351.

MARKOWITZ, H. M. (1952). Portfolio selection. *J. Finan.* **7**, 77–91.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press.

MCKEAN, J. W. and HETTMANSPERGER, T. P. (1978). A robust analysis of the general linear model based on one step *R*-estimates. *Biometrika* **65**, 571–579.

NAIR, U. S. (1936). The standard error of Gini's mean difference. *Biometrika* **28**, 428–436.

NYGÅRD, F. and A. SANDSTRÖM (1981). *Measuring Income Inequality.* Stockholm: Almqvist & Wiksell International.

OKUNEV, J. (1988). A comparative study of Gini mean difference and mean variance in portfolio analysis. *Acct. Finan.* **28**, 1–16.

PYATT, G., CHEN, CHAU-NAN and FEI, J. (1980). The distribution of income by factor components. *Quart. J. Econom.* **95**, 451–473.

RAO, V. M. (1969). Two decompositions of concentration ratio. *J. Roy. Statist. Soc. Ser. A* **132**, 418–425.

ROTHSCHILD, M. and STIGLITZ, J. E. (1970). Increasing risk: I. A definition. *J. Econom. Theory* **2**, 225–253.

ROY, A. D. (1953). Safety first and the holding of assets. *Econometrica*,

SANDSTRÖM, A. (1982). Various income sources and inequality of total income in Sweden in 1972 and 1976. *Statistisk Tidskrift* **4**, 288–299.

SCHECHTMAN, E. and YITZHAKI, S. (1987). A measure of association based on Gini's mean difference. *Comm. Statist. Theory Methods* **A16**, 207–231.

SEN, A. K. (1976). Poverty: An ordinal approach to measurement. *Econometrica* **44**, 219–231.

SHALIT, H. and YITZHAKI, S. (1984). Mean-Gini, portfolio theory and the pricing of risky assets. *J. Finan.* **39**, 1449–1468.

SHARPE, W. F. (1970). *Portfolio Theory and Capital Markets*, New York: McGraw Hill.

SHORROCKS, A. F. (1983a). The impact of income components on the distribution of family incomes. *Quart. J. Econom.* **98**, 311–326.

SHORROCKS, A. F. (1983b). Ranking income distributions. *Economica,* **50**, 3–17.

SIEVERS, G. L. (1983a). A robust multiple correlation coefficient for the rank analysis of linear models, Mathematics Report No. 69. Western Michigan University,

SIEVERS, G. L. (1983b). A weighted dispersion function for estimation in linear models. *Commun. Statist. Theory Methods* **12**, (10), 1161–1179.

SIMPSON, E. H. (1949). Measurement of diversity. *Nature*, 163, 688.

STUART, A. (1954). The correlation between variate values and ranks in samples from a continuous distribution. *British J. Statist. Psych.* **7**, 37–44.

SUITS, D. B. (1977). Measurement of tax progressivity. *Amer. Econom. Rev.* **67**, 747–752.

TAGUCHI, T. (1972a). On the two-dimensional concentration surface and extensions of the concentration coefficient and Pareto Distribution to the two dimensional case–I. *Ann. Inst. Statist. Math.* **24**, 355–382.

TAGUCHI, T. (1972b). On the two-dimensional concentration surface and extensions of the concentration coefficient and Pareto Distribution to the two dimensional case–II. *Ann. Inst. Statist. Math.* **24**, 599–619.

TAGUCHI, T. (1973). On the two-dimensional concentration surface and extensions of the concentration coefficient and Pareto Distribution to the two dimensional case–III. *Ann. Inst. Statist. Math.* **25**, 215–237.

YITZHAKI, S. (1982). Stochastic dominance, mean variance and Gini's mean difference. *Amer. Econom. Rev.* **72**, 178–185.

YITZHAKI, S. (1983). On an extension of the Gini inequality index. *Internat. Econom. Rev.* **24**, 617–618.

ZAGIER, D. (1983). Inequalities for the Gini coefficient of composite populations. *J. Math. Econom.* **12**, 103–118.

DEPARTMENT OF ECONOMICS
HEBREW UNIVERSITY
91905 JERUSALEM, ISRAEL

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305