METRON - *International Journal of Statistics*
2003, vol. LXI, n. 2, pp. 285-316

SHLOMO YITZHAKI

Gini's Mean difference: a superior measure of variability for non-normal distributions

Summary - Of all measures of variability, the variance is by far the most popular. This paper argues that Gini's Mean Difference (GMD), an alternative index of variability, shares many properties with the variance, but can be more informative about the properties of distributions that depart from normality. Its superiority over the variance is important whenever one is interested in one or more of the following properties: (a) stochastic dominance: the GMD can be used to form necessary conditions for stochastic dominance, while the variance cannot; (b) exchangeability: the GMD has two correlation coefficients associated with it, the difference between them being sensitive to the exchangeability between the marginal distributions. This property may play an important role whenever the "index number" problem is severe, (i.e., whenever the choice of the base for comparison between two marginal distributions may determine the direction of the results), or whenever the investigation procedure is based on an optimization procedure; (c) stratification: when the overall distribution is composed of sub-populations, the GMD is sensitive to stratification among sub-populations. The paper surveys the properties of the two indices of variability and discusses their relevance to several fields of research.

Key Words - Variance; Correlation; ANOVA; Gini Mean Difference.

1. Introduction

Of all measures of variability, the variance is by far the most popular. In this paper, it is argued that Gini's Mean Difference (Gini, 1912, hereafter GMD) is the most similar to the variance and shares many properties with it. However, while the variance is convenient and superior to the GMD for distributions that are nearly normal, the GMD is more informative about the underlined distributions. Its superiority over the variance is especially important in three fields: (a) Stochastic dominance: the GMD can be used to form necessary conditions for second-degree stochastic dominance, while the variance cannot. This property protects the investigator from making embar-

Received May 2003 and revised August 2003.

rassing mistakes in the fields of decision-making under risk and social welfare. (b) Exchangeability: it has two correlation coefficients associated with it, the difference between them is sensitive as to whether the marginal distributions are exchangeable. This property may play an important role in two types of cases. One arises when the ”index number” problem is severe, that is whenever the choice of the base for comparison may determine the direction of the results. For example, when measuring changes over time, the results may depend on the choice of whether we are moving forward or backward in time. Alternatively, when comparing countries then the results may differ according to the country chosen to be the base for comparison. The other arises whenever the investigation procedure is based on an optimization procedure. By construction, an optimization procedure results in an orthogonality condition which implies that a covariance (or covariances) is set to equal zero. Under the GMD framework the other covariance can be used to test the assumptions that led to the optimization. For example, in the case of regression analysis, using the GMD enables the investigator to construct a built-in specification test (Schechtman and Yitzhaki (2000)). (c) Stratification: when the overall distribution is composed of sub-populations, the GMD is sensitive to stratification among sub-populations. Therefore, we can claim that it reveals more information about the distributions.

Before making the case for the GMD, it is worthwhile to review the useful properties of the variance. The most important are:

(a) Normality: the variance is a parameter in the normal distribution. This is an advantage whenever the distribution is normal or converges to normal.

(b) Decomposability:
- (b.1) The variance of the sum of two random variables decomposes neatly into the sum of the variances, plus a term that is a function of the correlation between the two variables.
- (b.2) In the special case of two independent variables the variance of the sum is equal to the sum of the variances.
- (b.3) The variance of two mixed sub-populations can be decomposed into the intra and inter group variances.

(c) The variance is a quadratic function of the variate, and a quadratic function is easy to handle.

In contrast to the variance, the main advantage of the GMD lies in its non-decomposability. In general, it is impossible to decompose the GMD of a sum of two random variables into a formula that is similar to the decomposition of the variance. However, under certain conditions on the distributions it is possible to imitate the decomposition of the variance. The conditions that allow the imitation enable the investigator to observe additional properties of the distributions. Those properties may be important in areas like sociology and market structure. Another advantage of the GMD is that it is a member of a family of variability measures, the extended Gini family (Yitzhaki, 1983), that shares some of the main properties of the GMD, and it is useful for conducting sensitivity analysis. However, this family was mainly developed to fit economic theory and its adjustment to general data analysis brings complications. For this reason, I do not pursue this direction in this paper.

The structure of the paper is the following: Section 2 presents the similarities between the two measures of variability, Section 3 extends the comparison to the covariances and correlations associated with each index. Section 4 is devoted to the Absolute Concentration Curve, a tool that enables us to get additional information on the internal structure of the variance and GMD. Section 5 is devoted to the connection between Stochastic Dominance and GMD, Section 6 is devoted to the decomposition of the variability of the sum of random variables, while Section 7 deals with decomposition by population subgroups and stratification. Section 8 presents the extension to higher moments, and Section 9 discusses distances between distributions. Section 10 concludes.

To simplify the notation, I will make no distinction between population parameters and estimators. The estimator is always assumed to be the sample's version of the population parameter. However, there are two cases where one has to distinguish between the population and the sample. The population version will feature upper-case letters and be denoted with an equation number followed by a p. The sample version will be displayed with lower-case letters with an equation number followed by an s.

2. The similarities and differences between the variance and the GMD

2.1. Similarities

The most popular presentation of the variance is as a second central moment of the distribution. The most popular presentation of the GMD is as the expected absolute difference between two i.i.d. variables. (See Giorgi (1990) for a bibliographic portrait.) Using the absolute difference between two i.i.d. variables in order to measure variability characterized the Italian school, led by Corrado Gini, while reliance on central moments characterized the rest of the profession. This difference in outlook was the source of confrontation between the Italian school and what Gini viewed as the Western schools. (See Gini, 1965, 1966, p. 199; Hart, 1975.)

In retrospect, it seems that both sides were both right and wrong: both measures can be presented as central moments and also as based on the expected difference between two i.i.d. variables. To see this note the following:

2.1.1 –a)

Both indices can be defined without reference to a location parameter. Let $X_i (i = 1, 2)$ be two i.i.d. random variables. Then

$$G_X = E\{| X_1 - X_2 |\}, \tag{2.1}$$

where G_X is the GMD of X. In contrast

$$V_X = (1/2)E\{(X_1 - X_2)^2\}, \tag{2.2}$$

where V_x is the variance. That is, the GMD is the expected absolute difference between two randomly drawn observations, while the variance is the expected square of the same difference. It is interesting to note that Equation (2.2) raised to the power r is referred to as generalized mean difference (Gini, 1966, Ramasubban, 1958, 1959, 1960). However, as far as I know, they were not aware that when $r = 2$ it is identical to the variance. Furthermore, both indices can be written as a special case of a covariance:

$$G_X = 4\,\mathrm{COV}(X, F(X)), \tag{2.3}$$

whilst

$$V_X = \mathrm{COV}(X, X). \tag{2.4}$$

That is, the GMD is (four times) the covariance of a random variable with its cumulative distribution, $F(X)$, while the variance is the covariance of a random variable with itself (Yitzhaki, (1998). Stuart (1954) seems the first to identify this identity, for normal distributions).

As can be seen from (2.1) and (2.2), both indices can be defined without reference to a specific location parameter, such as the mean. This property can explain why it is possible, in regression analysis, to estimate the constant term in a different methodology then the one applied to estimate the slopes. To see the importance of this property note that one justification for Least Absolute Deviation regression (Bassett and Koenker, 1978) is that one may want the regression line to pass through the median (instead of the mean). If one wants to force the OLS (or GMD) regressions to pass through the median, one could first minimize the variance of the error term using version (2.2), in order to estimate the regression coefficients. Then one has to subtract the contribution of the independent variables from the dependent variable, and then to minimize the absolute deviation of the revised dependent variable by choosing a constant term. The result will be an OLS regression that passes through the median.

2.1.1 –b)

Both indices are sensitive to all observations. Note that the range, the absolute mean deviation, and inter-quantile difference do not have this property. The variance of the logarithm is sensitive to all observations, but it fails the Dalton principle of transfers. See Hart (1975, p.431). Foster and Ok (1999) show that it is not Lorenz-consistent.

2.1.1 –c)

Both indices can be presented graphically by the difference between the cumulative distribution and a first moment distribution, a curve which is known either as an Absolute Concentration (AC) curve or as a first moment distribution. This curve enables us to study the detailed structure of the variability index and the appropriate covariance that accompanies it, a property that is useful in determining the effect of a monotonic transformation on the variability index and on the sign of a regression coefficient. The presentation and implications of those properties will be discussed later following the description of AC curves.

2.1.1 –d)

Both indices can be presented as a weighted sum of distances between adjacent observations. In the population, the GMD can be written as:

$$\mathrm{COV}(X, F(X)) = 0.5 \int F(x)[1 - F(x)]dx\,. \tag{2.5.p}$$

To define the GMD in the sample, let $\Delta x_i, = x_{(i+1)} - x_{(i)}$, where $x_{(i)}$ is the order statistic, so that Δx_i, $i = 1, \ldots, n-1$ is the distance between adjacent observations. Then

$$\mathrm{cov}(x, F(x)) = \frac{1}{2n(n-1)} \sum_{i=1}^{n-1} (n-i)i\,\Delta x_i\,, \tag{2.5.s}$$

which means that the GMD attaches the largest weight to the section that is close to the median, and then, the weights declines symmetrically the farther is the section from the median.

The weighting scheme of the variance is given by:

$$\begin{aligned}\mathrm{cov}(x, x) &= \sum_{i=1}^{n-1}\sum_{j=i}^{n-1}\sum_{k=s}^{t-1}\sum_{p=s}^{t-1} \Delta x_k \Delta x_p \\ &= \sum_{i=1}^{n-1}\left(\sum_{j=i}^{n-1} i(n-j)\Delta x_j + \sum_{j=1}^{i-1} j(n-i)\Delta x_j\right)\Delta x_i\,,\end{aligned} \tag{2.6.s}$$

where $s = \min(i, j)$ and $t = \max(i, j)$ (Yitzhaki, 1996). The population version of (2.6.s) is (4.5), adjusted to represent the variance. To see the implication of (2.6.s), it is convenient to think in terms of an equi-distant distribution, so that $\Delta x_i = c$, for all i. In this case Equation (2.6.s) becomes

$$\mathrm{cov}(x, x) = K \sum_{i=1}^{n-1} (n-i) i c^2 , \tag{2.7.s}$$

where K is a normalizing factor, so that the weighting scheme is identical to the GMD weighting scheme. The only difference is that the GMD is a weighted average of c, while the variance is a weighted average of c^2 (Yitzhaki, 1996). Note that the range is the sum of Δx_i and therefore can be interpreted as assigning equal weight to each section Δx_i.

2.1.1 –e)

Both indices can be presented as a weighted sum of Order Statistics. To see this, rewrite the covariance formula of the GMD.

$$\mathrm{cov}(x, F(x)) = \frac{1}{n-1} \sum_{i=1}^{n} \frac{2i - (n+1)}{2n} x_{(i)} .$$

The variance can be written in a similar way.

2.2. Differences: City Block versus Euclidean

Both indices are based on averaging a function of the distance between two i.i.d. variables, a property that determines the weighting scheme illustrated in (2.5.s) and (2.6.s). In both cases the weighting scheme attaches the highest weight to the mid-rank observation (i.e., the median), and the weight declines symmetrically the farther the rank of the observation is from the mid-rank. The fundamental difference between the two measures of variability is the distance function they rely on. The GMD's distance function is referred to as the "city block" distance (or L1), while variance's distance is Euclidean. It is interesting to note that other measures of variability (e.g., the mean deviation) also rely on the L1 distance, but they do not share the weighting scheme caused by the averaging.

To shed some light on the difference in distance functions, note that the most basic measure of variability is the range, which is equivalent to the simple sum of the distances between adjacent observations, so that we end up with the difference between the most extreme parts of the distributions. If the distributions are restricted to have only two observations then the variance and the GMD (and all other measures of variability) will order all distributions in accordance with the ordering of the range. However, the range suffers from two major deficiencies: (1) it is not sensitive to the distribution of non-extreme observations, and (2) there are many important distributions with an infinite range. By comparing the distance between two i.i.d. variables, the range is decomposed into segments (distance between adjacent observations) and both indices have identical structure of weights. The difference is in the distance function used.

To illustrate the difference between the distance functions embodied in the GMD and the variance, we can ask, for a given range, which distribution has the smallest variance (GMD) and which has the maximum variance (GMD). Alternatively, for a given variance (GMD), we can ask which distribution has the largest range. Presumably, by answering those questions, we will be able to form an opinion as to which distance function is more appropriate, and which one reflects our intuition. To illustrate, restrict the distributions to three observations, and assume a given (normalized) range so that the discussion is restricted to distributions of the type: $[0, \delta, 1]$. Which δ maximizes or minimizes each variability index? Ignoring constants, the GMD is:

$$G(\delta) = \sum\sum \mid x_i - x_j \mid = 1 + \delta + \mid 1 - \delta \mid ,$$

so that the GMD equals 2 independent of the value of δ. Thus the position of the middle observation does not change the GMD. Repeating the same exercise with the variance yields

$$\sigma^2(\delta) = 1 + \delta^2 + (1 - \delta)^2 ,$$

so that the variance is maximized at $\delta = 0$ or 1, and minimized at $\delta = 0.5$. That is, the more equal are the distances defined by adjacent observations, the lower the variance. The result is that the variance is more sensitive than the GMD to variability in Δx_i, which is translated to sensitivity to extreme observations.

The alternative approach of evaluating the range for a given variability measure is presented geometrically. Figure 1 presents equal GMD and equal variance curves. On the horizontal axis is δ_1, (the difference between the second and first observations), while the vertical axis reports the difference between the third and second observations, δ_2. The range, of course is equal $\delta_1 + \delta_2$.

As can be seen, they represent different kind of distances: imagine you are in a city. If you are allowed to only move in the east/west or north/south directions then you are in a GMD (city block) world. If, on the other hand, you are allowed to move in any direction you want, and you are Pythagorean, then you are in a variance world. It is hard to determine which distance function should be preferred. However, the money metric, which is extensively used

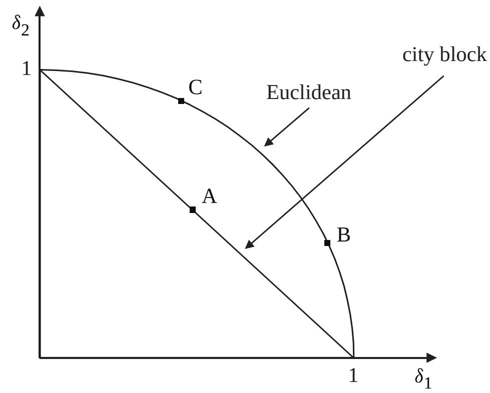

Figure 1. Equal GMD and variance curves.

by economists, resembles the city block metric because the distance function embodied in the budget constraint is identical to the distance function of the GMD (Deaton, 1979; McKenzie and Pearce, 1982; Jorgenson and Slesnick, 1984). Hence, one could have argued, that when it comes to choosing a metric economists have preferred the GMD type metric.

The implication of the difference in metrics can also be seen from the following question: consider the following distributions: [0, 0, 1] vs. [0, 0.575, 1.15]. Which distribution portrays a higher variability? If your intuitive answer points to the former (latter) distribution then you want to be in a variance (GMD) world (the variances are: 0.222, 0.220 respectively). This property is responsible for the sensitivity of Ordinary Least Squares regression to extreme observations. Another implication of the difference in metrics is that the GMD exists whenever the expected value exists while the existence of the variance requires the existence of a second moment.

The extension to more than 3 observations is identical. As can be seen from (2.1) and (2.2), both indices are based on differencing i.i.d. variables. Since the differencing is based on every two observations, the weight that is attached to distance between observations is greater the closer the observation to the median, as can be seen from (2.5.s) and (2.7.s). It is interesting to note that if the underlying distribution is normal, then the increase in the distance between adjacent observations when moving from the middle to the extremes is identical to the decrease in the weight due to being farther away from the median, so that each observation gets an equal weight (Yitzhaki, 1996). My conjecture is that this property leads to the statistical efficiency of variance-based statistics in cases of normality: the weights are distributed equally among observations.

3. Covariances and correlations

3.1.

The key parameters for decomposing the variance of the sums of random variables into its components are the covariances and the correlations, $\mathrm{COV}(X, Y)$ and $\rho = \mathrm{COV}(X, Y)/[\mathrm{COV}(X, X)\,\mathrm{COV}(Y, Y)]^{0.5}$. Equation (2.3) is suggestive of how to construct the covariance and correlation equivalents for the GMD. The GMD forms two asymmetric correlation coefficients, between two random variables. Let (X, Y) be two random variables. Then:

$$\mathrm{COV}(Y, F_X(X)) \quad \text{and} \quad \mathrm{COV}(X, F_Y(Y))\,, \tag{3.1}$$

are the two covariance-equivalents associated with the GMD, and

$$\Gamma_{XY} = \frac{\mathrm{COV}(X, F_Y(Y))}{\mathrm{COV}(X, F_X(X))} \quad \text{and} \quad \Gamma_{YX} = \frac{\mathrm{COV}(Y, F_X(X))}{\mathrm{COV}(Y, F_Y(Y))}\,, \tag{3.2}$$

are the (asymmetric) Gini correlation coefficients. Blitz and Brittain (1964) defined the Gini correlation by using concentration and Lorenz curves, while Kendall (1955) mentions it, but without connecting it to the GMD. The properties of Gini correlation (G-correlation hereafter) coefficients are listed in Schechtman and Yitzhaki (1987, 1999) and Yitzhaki and Olkin (1991):

(a) The G-correlation coefficient is bounded, such that $1 \geq \Gamma_{js} \geq -1(j, s = X, Y)$; an important property of the G-correlation is that the bounds are identical for all marginal distributions. (Schechtman and Yitzhaki, 1999). Note, that this property does not hold for the Pearson's correlation coefficient (De Veaux, 1976; Kotz and Seeger, 1991; Shih and Huang, 1992).
(b) If X and Y are independent, then $\Gamma_{YX} = \Gamma_{XY} = 0$;
(c) Γ_{YX} is not sensitive to a monotonic transformation of X; (similar to Spearman's correlation coefficient).
(d) Γ_{YX} is not sensitive to a linear monotonic transformation of Y; (similar to Pearson's correlation coefficient).
(e) In general, Γ_{js} need not be equal to Γ_{sj} and they may even have different signs. However, if the random variables Z_j and Z_s are exchangeable up to a linear transformation, then $\Gamma_{js} = \Gamma_{sj}$. Exchangeability implies that the shape of the marginal distributions is similar, and that the relationships between the variables are symmetric. Formally, as defined by Stuart and Ord (1987: 419), "*a set of random variables is said to be exchangeable if, for all n* ≥ 1, *and for every permutation of n subscripts, the joint distributions of* $(x_{j1}, \ldots, x_{jn})$ *are identical.*"

The importance of property (a), which states that the G-correlations are always bounded by the same bound, can be illustrated by comparison to a

Pearson correlation coefficient of 0.39. Is this a high correlation? Knowing that the upper bound is 0.4 can change one's impression of the association between the variables.

Property (e), symmetry under exchangeability up to a linear transformation, enables the decomposition of the GMD of a sum of random variables in a way which is identical to the decomposition of the variance. We will return to this property in Section 6.

In some cases, one may be interested in imposing a symmetric version of the correlation coefficients, i.e., to impose the property $S_{XY} = S_{YX}$, for any two distributions, as is the case in the Pearson's and Spearman's correlation coefficients. There are several ways of defining a symmetric version of G-correlations (Yitzhaki and Olkin, 1991), but the most useful one is:

$$S = \frac{G_X \Gamma_{XY} + G_Y \Gamma_{YX}}{G_X + G_Y}\,, \tag{3.3}$$

so that the symmetric correlation coefficient is a weighted average of the asymmetric ones, weighted by the variability index (Yitzhaki and Wodon, 2000). The properties of (3.3) are discussed in Section 6.

3.2.

Both the Gini covariance (hereafter co-Gini) and the regular covariance can be presented by concentration curves. The concentration curve enables one to see whether the relationship between the variables is monotonic or whether it may differ conditional on the section of the distribution examined. The discussion of this property will be carried out in Section 4.

3.3.

The property of having two covariances for two random variables is an extremely important one in cases in which the results are derived by an optimization procedure, as is the case of OLS regression. In the OLS example, the optimization requires setting the covariance between the independent variable and the error term equal to zero (the normal equation). Having two covariances implies that although one covariance is set to zero, the other covariance is free to have any value. Hence, one should be required to test whether the other covariance is also equal to zero. Rejection of the hypothesis that the other covariance is equal to zero implies rejection of the validity of the assumptions that led to the optimization. Therefore, unlike the OLS, which is based on the variance, Gini-regressions have two versions, depending on which co-Gini is set to zero, and the existence of the other covariance enables testing the linearity of the regression. (See, Olkin and Yitzhaki, 1992 for the simple regression case and Schechtman and Yitzhaki, 2000 for an extension to multiple regression).

Note that the two covariances defined by the GMD, may have different signs. This property implies that the Gini regression coefficient of X on Y may have a different sign than that of the regression coefficient of Y on X. Such a situation cannot occur in the OLS, because $\mathrm{COV}(X, Y) = \mathrm{COV}(Y, X)$. Further research is needed to evaluate the meaning and implications of such an observation.

3.4.

The similarities between the Pearson and Gini correlations can be best shown and investigated by using the copula. (See Nelsen (1999) for definitions and properties of the copula, and Rockinger and Jondeau (2001) for an application in finance). Let $D(X, Y) = F_{XY}(X, Y) - F_X(X)F_Y(Y)$, where $F_{XY}(X, Y)$ is the joint cumulative distribution, be the deviation of the joint cumulative distribution from independence. Then it is shown by Schechtman and Yitzhaki (1999) that

$$\rho = \frac{1}{\sigma_X \sigma_Y} \iint D(x, y) dx dy \, ,$$

where ρ is Pearson's correlation coefficient, while

$$\Gamma_{XY} = \frac{1}{G_X} \iint D(x, y) dx dF_Y(y) \, .$$

Thus, the only difference between Pearson's and GMD correlation is in whether the integration is done over dy or $dF_Y(y)$. This leads us to the conjecture that any structure imposed on $D(X, Y)$ can be translated to GMD or Pearson correlation in a similar way.

4. The absolute concentration curve

The concentration curve is a generalization of the well-known Lorenz curve. It is widely used in the field of income distribution to portray the impact of taxes on the distribution of real income or to estimate the income elasticity of a consumption good (Kakwani, 1977, 1980; Lambert, 1993; Suits, 1977; and Yitzhaki and Slemrod, 1991). In a typical case, the horizontal axis would portray the poorest p percent of the population while the vertical axis would present the share of total expenditure on a consumption item spent by the poorest p percent. The absolute concentration curve (ACC) differs from the concentration curve by presenting the *cumulative consumption* (rather than the cumulative share of consumption) of the poorest p percent on the vertical axis. In this paper I am interested in the ACC as a general tool that can be applied to any random variable. The definitions follow the terminology in Yitzhaki and Olkin (1991). The conclusion I would like to draw from this section is that the parameters of both approaches can be geometrically presented by versions of the concentration curves. However, the curves representing the GMD parameters are bounded and therefore have a simple structure.

The following notation is used. A bivariate density function is denoted by $f_{XY}(x, y)$, or (when no confusion arises) more simply by $f(x, y)$. Marginal distributions are denoted $f_X(x)$ and $f_Y(y)$. The means, or expected values, are μ_X and μ_Y. The conditional density function is $f_{Y|X}$ and the conditional expectation is $g(x) = \mu_{Y.X} \equiv E_Y(Y \mid X = x)$. It is assumed that all densities are continuous and differentiable, and all second moments exist.

Definition 4.1. The *absolute concentration curve* (ACC) of Y with respect to X, denoted $A_{Y.X}(p)$, is implicitly defined by the relationship

$$A_{Y.X}(p) = \int_{-\infty}^{X_p} g(t)dF_X(t)\,, \tag{4.1}$$

where X_p is defined by

$$p = \int_{-\infty}^{X_p} dF_X(t)\,. \tag{4.2}$$

A special case of the absolute concentration curve is $A_{X.X}(p)$. We will refer to it as the *absolute Lorenz curve*. Shorrocks (1983b) calls this curve the Generalized Lorenz Curve. Gastwirth (1971) presents a formal definition of the Lorenz curve (see also Pietra (1915, 1931).

The following definitions will be needed later:

Definition 4.2. The Line of Independence (LOI) is the line connecting $(0, 0)$ with $(\mu_Y, 1)$. Let $L_{Y.X}(p) = \mu_Y p$ denote the LOI of Y with respect to X.

Figure 2 presents an ACC curve and LOI that are convenient for our purposes. The solid curve $[0ABCDE]$ is the absolute concentration curve of Y with respect to X and the dashed line $[0FCE]$ is the LOI.

To apply the concentration curve to variance-related parameters, e.g., ordinary least squares regression, it is convenient to redefine the concentration curve and the LOI as functions of the variate, X. In this case, we denote the transformed ACC by VA and define it as

$$\mathrm{VA}_{Y.X}(x) = \int_{-\infty}^{X} g(t)dF_X(t)\,. \tag{4.3}$$

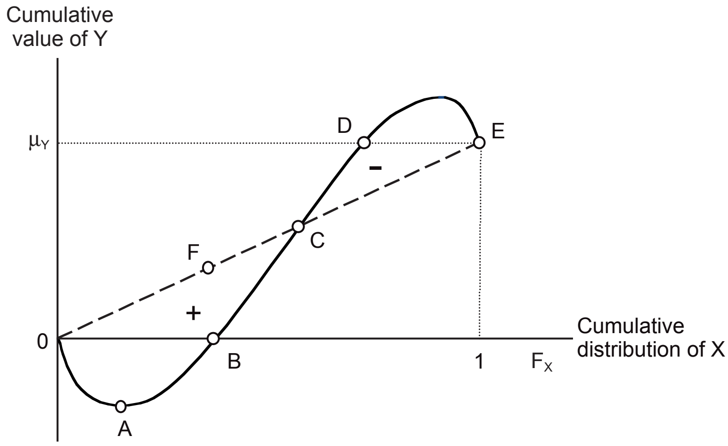

Figure 2. The absolute concentration curve.

The LOI simply changes to $L_{Y.X}(p(x)) = \mu_Y F_X(x)$. Note, however, that it is no longer a straight line. In terms of figure 2 the only difference between equations (4.1) and (4.3) is that the horizontal axis is changed from $p = F_X(x)$ to x.

The ACC of Y with respect to X has the following properties (Unless specifically stated the listed properties are for A and not VA curve):

(a) It passes through the points $(0, 0)$ and $(\mu_Y, 1)$.
(b) Its derivative with respect to p is $E_Y(Y \mid X = X_P)$. This follows directly from Definition 4.1. Consequently, $A_{Y.X}(p)$ is increasing if and only if $g(x) = E_Y(Y \mid X = X_P) > 0$.
(c) It is convex (concave, straight line) if and only if $\partial g(x)/\partial x > 0$. ($\partial g(x)/\partial x < 0$, $\partial g(x)/\partial x = 0$). $A_{X.X}(p)$ is always convex.

If Y and X are independent, the ACC coincides with the LOI.

(1) The area between the LOI and the ACC is equal to $\mathrm{COV}(Y, F_X(X))$, that is,

$$\mathrm{COV}(Y, F_X(X)) = \int_0^1 \{\mu_Y p - A_{Y.X}(p)\} dp \,. \qquad (4.4)$$

The term $\mathrm{COV}(Y, F_X(X))$, i.e., the covariance between a random variable and the cumulative distribution of another variable is the equivalent of the regular covariance when the Gini mean difference (GMD) is used as a measure of variability.

(2) The area between the shifted (transformed) LOI and the shifted ACC is equal to $\mathrm{COV}(Y, X)$. That is:

$$\mathrm{COV}(Y, X) = \int_{-\infty}^{\infty} \{\mu_Y F_X(t) - \mathrm{VA}_{Y.X}(t)\} dt\,. \tag{4.5}$$

Note that $\mathrm{VA}_{Y.X}(x)$ is the transformed ACC, while $\mu_Y F_X(t)$ is the transformed LOI. The variance of X is the area between the (transformed) LOI and the (transformed) ACC, denoted by $\mathrm{VA}_{X.X}$ (see Yitzhaki, 1998).

Property (f) below is a modification of a result in Grether (1974), who pointed out that, in general, the signs of Pearson's correlation coefficients are not invariant to order-preserving transformations of one of the variables.

(f) The ACC is above the LOI for all p if and only if $\mathrm{COV}(Y, T(X)) < 0$ for any continuous differentiable monotonically increasing function $T(X)$. (The ACC is below the LOI iff the covariance is positive.)

This implies that whenever the ACC intersects the LOI one can divide the data into two sections, conditional on the values of X. In one section $\mathrm{COV}(Y, X) < 0$ and in the other $\mathrm{COV}(Y, X) > 0$. By applying a monotonic transformation to X, the investigator can change the magnitude of the covariances in those sections, thereby affecting the sign of the overall covariance. This property is used in Yitzhaki (1990) to show the conditions under which it is possible to change the sign of a regression coefficient by applying a monotonic transformation to one of the variables. To illustrate, the ACC in figure 2 intersects the LOI at C. Property (f) implies that the data used in figure 2 can be divided into two groups according to whether X is greater or lower than X_C, the limit defined by C. The group $(X < X_C)$ would yield a positive covariance between Y and any monotonic transformation of X, while the group $(X > X_C)$ will show a negative covariance. A transformation can increase (decrease) the absolute value of a covariance if $T'(x) > 1 (< 1)$. Note, however, that if $A_{Y.X}$ and $L_{Y.X}$ intersect, it does not necessarily imply that $A_{X.Y}$ and $L_{X.Y}$ intersect.

(g) If, as is sometimes assumed, Y and X follow a bi-variate normal distribution, $A_{Y.X}$ and $L_{Y.X}$ do not intersect. This means that a monotonic transformation of X or Y cannot change the sign of the covariance.
(h) The maximum vertical distance between $A_{Y.X}$ and $L_{Y.X}$ is bounded from above by the absolute deviation of Y. The proof is trivial, following properties (b) and (c).

Additional properties of ACC are presented in Yitzhaki and Olkin (1988).

The estimators of $L_{Y.X}$ and $A_{Y.X}$ are based on what is known in statistics as concomitant Y's of X (that is, the observations of Y's, after ordering them by their respective X values. See Barnett, Green and Robinson (1976) for properties of concomitants). Then:

$$L_{Y.X}(p) = p\overline{y}_n\,, \tag{4.6}$$

where $p = i/n$ and $\overline{y}_n$ is the mean of Y. The empirical ACC is

$$A_{Y.X}(p) = \frac{1}{n}\sum_{j=1}^{i} y_j = p\overline{y}_i \, , \tag{4.7}$$

where $\overline{y}_i$ is the mean of the first i observations of Y, ordered according to their respective X values. The vertical difference between the LOI and ACC is:

$$L_{Y.X}(p) - A_{Y.X}(p) = p(\overline{y}_n - \overline{y}_i) \, , \tag{4.8}$$

where $p = i/n$.

Finally, there are several theorems concerning the convergence of empirical concentration curves to the population curves (Goldie, 1977; Bishop, Formby and Thistle, 1992), and several large sample tests concerning the intersection of concentration curves. (Eubank, Schechtman and Yitzhaki, 1993; Beach and Davidson, 1983; Bishop, Chow and Formby, 1994; Davidson and Duclos, 1997; and Nygârd, F. and A. Sandström, 1981). However, convergence issues and estimation issues (Gastwirth (1972), Hoeffding (1948)) are beyond the scope of this paper.

5. GMD AND SECOND DEGREE STOCHASTIC DOMINANCE

One of the dominant approaches to modeling behavior under risk is the expected utility hypothesis. According to this approach, the decision maker maximizes the expected value of his utility function, defined over wealth or income. Since the utility function is not observed, it is important to reduce the assumptions concerning the shape of the utility function, because the less one assumes the more general is the analysis. Under the second degree stochastic dominance paradigm, the only assumptions about the utility function is that the first derivative is positive, and the second derivative is negative (Hadar and Russell, 1969; Hanoch and Levy, 1969). See the survey by Levy (1992). The implementation of the methodology offers only a partial ordering over the set of risky options. Furthermore, as far as I know, no algorithm has been found that is able to apply it in the typical portfolio allocation framework, which is characterized by several risky assets. Therefore, it is convenient to use two summary statistics to describe each risky option, the mean of the distribution, which represents the reward, and a variability index that represents the risk. The two-parameter summary statistics approach offers scope for a complete ordering, and can handle multiple risky assets. However, it is important to verify that the use of summary statistics does not contradict the expected utility hypothesis.

A popular approach is to use the mean, μ, and the variance, σ^2, as the summary statistics (MV approach) and to define the utility function over the summary statistics, $U(\mu, \sigma)$ with the derivative $U_\mu > 0$, and $U_V < 0$. This approach produces identical ranking to the expected utility approach if the distributions are restricted to belong to the multivariate normal, or if the utility function is quadratic (See Feldstein, 1969; Tobin, 1958, 1965; Hanoch and Levy, 1969; Levy and Markowitz, 1979). An alternative approach is to use the GMD instead of the variance (Yitzhaki, 1982; Shalit and Yitzhaki, 1984), the MG approach. Unlike the MV, the mean and the GMD can be used to construct necessary conditions for second degree stochastic dominance:

Proposition 1. $\mu_X \geq \mu_Y$ *and* $\mu_X - 2\,\mathrm{cov}(X, F_X(X)) \geq \mu_Y - 2\,\mathrm{cov}(Y, F_Y(Y))$ *are necessary conditions for* $E_X\{U(X)\} \geq E_Y\{U(Y)\}$, *for all U with* $U' \geq 0$, *and* $U'' \leq 0$.

If the cumulative distributions of X and Y are restricted to cross at most once, then the above conditions are sufficient as well.

Proof. Yitzhaki (1982). If it is assumed that $U'' > 0$, then the second inequality translates into $\mu_X + 2\,\mathrm{cov}(X, F_X(X)) \geq \mu_Y + 2\,\mathrm{cov}(Y, F_Y(Y))$ (Yitzhaki, 1999). However, most of the interest of economists is limited to concave functions, where $U'' < 0$.

To see why this property is important, consider using a complicated computer model to evaluate the following two distributions using the mean-variance model: one distribution is restricted to [0,1] while the other is on [1000, 2000]. The mean and the variance of the latter distribution are greater than the mean and the variance of the former. This leads us to the embarrassing conclusion that it is impossible to conclude which distribution is the preferred one. By forming necessary conditions for stochastic dominance, the investigator is able to omit from the mean-variance (or MG) efficient set these portfolios that are (second-degree) stochastically dominated by other portfolios. Proposition 1 enables one to restrict the MG efficient set to a subset of the Second-degree stochastic dominance efficient set. Similar conditions can be constructed using the variance, but they are restricted to distributions with equal means, or to normal distributions. (See, Formby, Smith and Zheng, 1999).

6. The decomposition of the variability of a sum of variables

The decomposition of the variance of a linear combination of random variables is a basic tool used in regression analysis, portfolio analysis and many other empirical applications. Hence, it is important to see whether the

GMD has similar properties.

Let $Z = \alpha X + (1-\alpha)Y$. Then:

$$\sigma_Z^2 = \alpha\sigma_X^2 + (1-\alpha)^2\sigma_Y^2 + 2\alpha(1-\alpha)\sigma_X\sigma_Y\rho_{XY} , \tag{6.1}$$

where ρ is Pearson's correlation coefficient. The GMD of Z can be decomposed in a similar way, provided that some conditions on the underlying distribution hold. The violation of those conditions adds additional terms to the decomposition.

Proposition 2. *Let (X, Y) follow a bi-variate distribution. Let $Z = \alpha X + (1-\alpha)Y$, where α is a given constant. Then*

(a)

$$\begin{aligned} &G_Z^2 - [\alpha D_{XZ}G_X + (1-\alpha)D_{YZ}G_Y]G_Z \\ &\quad = \alpha^2 G_X^2 + (1-\alpha)^2 G_Y^2 + \alpha(1-\alpha)G_X G_Y(\Gamma_{XY} + \Gamma_{YX}) \end{aligned} \tag{6.2}$$

where $G_Y = \mathrm{COV}(Y, F(Y))$ is (one-fourth) of Gini's mean difference, $\Gamma_{ZY} = \frac{\mathrm{COV}(Z,F(Y))}{\mathrm{COV}(Z,F(Z))}$ is the Gini correlation, and $D_{ij} = \Gamma_{ij} - \Gamma_{ji}$ $(i, j = X, Y, Z)$ is the difference between the Gini correlations.

(b) *Provided that $D_{ij} = 0$, for all X, Y, Z then the following decomposition holds:*

$$G_Z^2 = \alpha^2 G_X^2 + (1-\alpha)^2 G_Y^2 + 2\alpha(1-\alpha)G_1 G_2 \Gamma , \tag{6.3}$$

where $\Gamma = \Gamma_{XY} = \Gamma_{YX}$ is the Gini correlation between X and Y.

Clearly, case (6.3) is a special case of (6.2). Note that the right hand side of (6.2) and (6.3) is identical to the decomposition of the variance of a weighted sum of random variables, except the the Gini parameters substitute for the variance parameters. Because of its similarity to the variance decomposition, the practical importance of case (6.3) is much higher than that of case (6.2) because it implies that any variance-based model can be replicated, using the GMD as a substitute for the variance. On the other hand, case (6.3) offers additional information about the underlying distribution. Further research is needed to evaluate the usefulness of this additional information.

Proof of (6.2). Using the properties of the covariance we can write:

$$\begin{aligned} G_Z &= \mathrm{COV}[\alpha X + (1-\alpha)Y, F(Z)] \\ &= \alpha\,\mathrm{COV}[X, F(Z)] + (1-\alpha)\,\mathrm{COV}[Y, F(Z)] \\ &= \alpha\Gamma_{XZ}G_X + (1-\alpha)\Gamma_{YZ}G_Y . \end{aligned} \tag{6.4}$$

Define now, the identity:

$$\Gamma_{ij} = D_{ij} + \Gamma_{ji} \quad \text{for} \quad i = X, Y, Z .$$

Where D_{ij} is the difference between the two Gini correlations. Using the identity, we get:

$$G_Z = \alpha(\Gamma_{ZX} + D_{ZX})G_X + (1-\alpha)(\Gamma_{ZY} + D_{ZY})G_Y$$

Rearranging terms:

$$G_Z - \alpha D_{XZ}G_X - (1-\alpha)D_{YZ}G_Y = \alpha\Gamma_{ZX}G_X + (1-\alpha)\Gamma_{ZY}G_Y$$

But, using the properties of the covariance:

$$\begin{aligned}\Gamma_{ZX} &= \frac{\operatorname{cov}(Z, F(X))}{\operatorname{cov}(Z, F(Z))} = \frac{1}{\operatorname{cov}(Z, F(Z))}\\ &\quad \times \{\alpha \operatorname{cov}(X, F(X)) + (1-\alpha)\operatorname{cov}(Y, F(X))\}\\ &= \frac{\alpha G_X + (1-\alpha)G_Y\Gamma_{YX}}{G_Z}\,.\end{aligned}$$

Writing Γ_{ZY} in a similar manner, and applying it to (6.4) we get:

$$\begin{aligned}&G_Z^2 - [\alpha D_{XZ}G_X + (1-\alpha)D_{YZ}G_Y]G_Z\\ &\quad = \alpha G_X(\alpha G_X + (1-\alpha)G_Y\Gamma_{YX}) + (1-\alpha)G_Y(\alpha\Gamma_{XY}G_X + (1-\alpha)G_Y)\\ &\quad = \alpha^2 G_X^2 + (1-\alpha)^2 G_Y^2 + \alpha(1-\alpha)G_X G_Y(\Gamma_{XY} + \Gamma_{YX})\,. \qquad \square\end{aligned}$$

Proof of (6.3). Assuming equality of the Gini correlation coefficients between Z and X implies that $D_{XZ} = 0$. A similar assumption with respect to violation of the assumption of equality between Z and Y correlations implies $D_{YZ} = 0$. The assumption of $\Gamma = \Gamma_{XY} = \Gamma_{YX}$ completes the proof of (b). □

Clearly, equation (6.3) is easier to work with than (6.2). The question arises is how restricting are the assumptions $D_{ij} = 0$, $i, j = X, Y, Z$. Schechtman and Yitzhaki (1987) show that a sufficient condition for $D_{ij} = 0$ is that the variables are exchangeable up to a linear transformation.

The necessary and sufficient conditions for equality of Gini correlations can be stated in terms of concentration curves. To take into account the "up to a linear transformation" requirement, let X^N, Y^N be the normalized variables:$X^N = (X - \mu_X)/G_X$ where μ is the expected value and $G_X = \mathrm{COV}(X, F(X))$, and $Y^N = (Y - \mu_Y)/G_Y$. The variables X^N and Y^N have expected value of zero and a GMD of one. Since the Gini correlations are not sensitive to a linear monotonic transformation of the variables, $\Gamma_{XY} = \Gamma_{X^N Y^N}$ and $\Gamma_{YX} = \Gamma_{Y^N X^N}$, so that equality of the correlations of the normalized variables implies equality of the Gini correlations of the original variables.

Let $e_X(t) = E\{Y^N \mid X^N = t\}$ and $e_Y(t) = E\{X^N \mid Y^N = t\}$ be the conditional expectations.

Proposition 3. *A necessary and sufficient condition for* $\Gamma_{XY} = \Gamma_{YX}$ *is*:

$$\int e_{Y^N}(Y^N(p))pdp = \int e_{X^N}(X^N(p))pdp\,,$$

where $Y^N(p)$ *and* $X^N(p)$ *are the inverse of the marginal distribution of* Y^N *and* X^N, *respectively.*

Proposition 3 has a simple geometric presentation in terms of absolute concentration curves. It implies that for the two Gini correlations to be equal it is necessary and sufficient that the area enclosed by the concentration curve of X^N with respect to Y^N is equal to the area enclosed by the concentration curve of Y^N with respect to X^N.

Proof. The denominator in a Gini correlation of the normalized random variables is equal to one by construction. Hence,

$$\begin{aligned}\Gamma_{XY} &= \mathrm{COV}(X^N, F_{Y^N}(Y^N)) = E\{e_{Y^N}(Y^N)F_{Y^N}(Y^N)\}\\ &= \int_{-\infty}^{\infty} e_{Y^N}(Y^N)F_{Y^N}(Y^N)dF_{Y^N}(Y^N) = \int_0^1 e_{Y^N}(Y^N(p))pdp\,,\end{aligned} \tag{6.5}$$

where $Y^N(p)$ is the inverse of the marginal distribution of Y^N. The first equality in (6.5) holds because the expected value of the normalized variable equals zero. The last equality is due to transformation of variable $p = F()$. The same procedure can be applied to Γ_{YX}. It can be immediately seen that $e_X(X^N(p)) = e_Y(Y^N(p))$ for all p is a sufficient condition for $\Gamma_{XY} = \Gamma_{YX}$ while equality of the integrals is a necessary and sufficient condition. □

To illustrate the interpretation of condition (6.5), it is useful to present it in terms of concentration curves. Define the absolute concentration curve as:

$$A_{X.Y}(p) = \int_0^p e_Y(Y(t))dt\,, \tag{6.6}$$

with the cumulative value of X given $F_Y(Y) \le p$. Then:

$$\int_0^1 e_{Y^N}(Y^N(p))pdp = \int_0^1 \frac{\partial A_{X^N.Y^N}(p)}{\partial p}pdp = -\int_0^1 A_{X^N.Y^N}(p)dp\,. \tag{6.7}$$

where the second step is derived by integration by parts, using the property that the expected value of X^N is zero.

Note that requiring $\Gamma_{XY} = \Gamma_{YX}$ to hold, imposes no restriction on the shape of $A_{X.Y}$ versus the shape of $A_{Y.X}$. In particular, the absolute concentration curves are allowed to cross the horizontal axis, which means that the conditional correlations in segments of the distribution can be of different signs. It allows the conditional distributions to be of totally different shape, and hence they do not have to belong to the same family, nor does the joint distribution need to be symmetric.

Assuming a multivariate distribution of a specific family impose strict conditions on the concentration curves, so they are not only identical but also change in an identical manner along the change in parameters of the distribution. To illustrate the kind of conditions imposed by the multi-variate normal distribution, note that in this case $A_{Y^N X^N}(p) = A_{X^N Y^N}(p) = \rho A_{X^N X^N}(p) = \rho A_{Y^N Y^N}(p)$ for all $0 < p < 1$, where ρ is Pearson's (or Gini) correlation coefficient.

This implies that the sufficient condition for equality of the Gini correlations is much weaker than exchangeability, and the necessary and sufficient conditions can be tested by comparing concentration curves. Further research is needed to evaluate the implications of the additional terms of the GMD decomposition.

Finally, the extension of (6.2) to more than 2 variables is immediate. It is given without a proof. Let $Y_0 = \sum_{i=1}^{k} a_i Y_i$, then

$$G_0^2 - G_0 \sum_{i=1}^{k} a_i D_{i0} G_i = \sum_{i=1}^{k} a_i^2 G_i^2 + \sum_{i=1} \sum_{i \neq j} a_i a_j G_i G_j \Gamma_{ij} \tag{6.8}$$

The additional terms in a GMD decomposition can play an important role in a convergence to normality process. Consider a sequence of i.i.d. variables and let $Y_k = (X_1 +, \dots, +X_k)/k$. Then the distribution of Y_k converges to the normal distribution. However, if the distribution is normal the additional terms in the GMD decomposition should be equal to zero. Hence, whenever we are dealing with a convergence process, we can split the sample into two independent samples. In each sub-sample we take the averages of k observations. If aggregation to size k implies that we have already converged to a normal distribution then we should expect the GMD decomposition of the overall sample to two samples of averages of k observations to behave like the decomposition of the variance. The value of the additional terms can be used to evaluate the stage of the convergence.

7. The decomposition of a variability index into sub-groups

It is well known that the variance of the union of several sub-populations can be neatly decomposed into intra- and inter-group components, with the inter-group component being the variance of the means appropriately weighted, and the intra-group component being the weighted sum of the internal variances (ANOVA). This result holds for any distribution.

In contrast, the GMD does not decompose neatly into two components. Using the covariance presentation of the GMD (equation 2.3), each observation is represented by its rank and by the variate. Similarly, in the GMD framework, the group should be represented by the mean rank of its members in the overall distribution, and the mean variate. Using this representation, one can decompose the GMD into a formula in the following way.

Let Y_i, $F_i(Y)$, $f_i(Y)$, μ_i, p_i, represent the random variable, the cumulative distribution, the density function, the expected value, and the share of group i in the overall population, respectively. The overall population, $Y_o = Y_1 \cup Y_2, \ldots, \cup Y_n$ is the union of all groups. It is shown in Yitzhaki (1994) that:

$$G_o = \sum_{i=1}^{n} p_i O_i G_i + GB \,, \tag{7.1}$$

where G_i is the GMD of group i, O_i is an index of the overlapping of group i with the overall population, (defined below), and GB is the between-group GMD (also defined below). Before proceeding, it is worth mentioning that in the area of income distribution, most interest focuses on the Gini coefficient, which is a relative measure of variability (similar to the coefficient of variation). The decomposition presented here is of the GMD. The decomposition of the Gini coefficient follows similar lines.

Using these concepts the GMD can be decomposed similarly to the variance, provided that the distributions of the groups are perfectly stratified. Alternative decompositions also exist but they will not be pursued in this paper (Rao (1969), Dagum (1987), Silber (1989)).

7.1. The overlapping index

The concept of overlapping is closely related to two other concepts frequently used in social sciences – stratification and segmentation. Stratification is defined by Lasswell (1965, p. 10) as follows:

"In its general meaning, a stratum is a horizontal layer, usually thought of as between, above or below other such layers or strata. Stratification is the process of forming observable layers, or the state of being comprised of layers. Social stratification suggests a model in which the mass of society is constructed of layer upon layer of congealed population qualities."

According to Lasswell, perfect stratification occurs when the observations of each group are confined to a specific range, and the ranges of different groups do not overlap. An increase in overlapping can therefore be interpreted as a reduction in stratification. Stratification plays an important role in the theory of relative deprivation (Runciman (1966)), who argues that stratified societies can tolerate greater inequality (see also Yitzhaki (1982)).

Stratification may also be important in other areas of economics. Consider a product that can be hierarchically ranked according to one attribute. For example, assume that cars can be ranked according to price. Consider now several producers, each producing several types of cars. We may define a market as (perfectly) stratified if each producer offers cars in a certain price range, and those price ranges are separated, so that no producer offers a car in a price range of another producer. On the other hand, the market will be (perfectly) non-stratified if all producers offer the same range of prices for cars. Market segmentation is of interest because the less segmented is the market, the more competitive it is likely to be.

Another example of stratification concerns school segregation (segregation may be interpreted as stratification without the hierarchical connotation, e.g., boys and girls). Consider a population of students that can be ranked according to ability. The distribution of students will be perfectly stratified by schools if the students of each school occupy a certain range in the ability distribution and no student from another school is in that range. Stratification of the students population by schools allows one to predict students' performance from knowing their school. Also, as shown in Yitzhaki and Eisenstaedt (2003), overlapping of distributions is an important factor in determining whether one can find an alternative examination that will reverse the order of average scores of students in different schools. Stratification is a property that may be negatively correlated with variability, but need not be identified with it. Hence, it is important to separate the two in a decomposition framework, so that one can distinguish stratification from non-variability.

In the real world one can rarely identify situations in which there is no overlapping. We must therefore define situations of near-perfect stratification and an index of the "degree of overlapping."

Definition. Distribution j is perfectly overlapped by distribution i if the observations of distribution j are either below or above the observations of distribution i.

We will say that distribution i out-spans distribution j. The overlapping index quantifies the term "out-span".

The overlapping index for group i, with respect to group j, is defined as:

$$O_{ji} = \mathrm{COV}_i[Y, F_j(Y)]/\,\mathrm{COV}_i[Y, F_i(Y)]\,, \tag{7.2}$$

where COV_i denotes that the density function used in calculating the covariance is that of distribution i. Specifically,

$$\mathrm{cov}_i[Y, F_j(Y)] = \int (y - \mu_i)(F_j(y) - F_{ji})dF_i(y)\,, \tag{7.3}$$

and

$$F_{ji} = \int F_j(t)dF_i(t) \tag{7.4}$$

is the expected rank of observations of distribution i had they been ranked according to the ranking of group j. (In the sample, the cumulative distribution is estimated by the rank of the observation.) Alternatively, the overlapping index is the G-correlation of observations of group i, with their assigned rank, had they been ranked according to distribution j. Note that ranking observations according to various reference groups is common in sports, where athletes are ranked according to different groups (e.g., national, world,etc ...) and in banking, where banks are ranked according to national and world ranking.

The overlapping index measures the degree to which distribution j is included in the range of distribution i. Its properties are the following:

(a) $O_{ji} \geq 0$. The index is equal to zero if no member of the j distribution is in the range of distribution i.
(b) The index is an increasing function of the fraction of group j that is located in the range of group i.
(c) For a given fraction of distribution j that is in the range of distribution i, the smaller the distance between the observations belonging to j to the mean of group i, the higher is O_{ji}.
(d) If the distribution of group j is identical to the distribution of group i, then $O_{ji} = 1$. Note that by definition $O_{ii} = 1$.
(e) $O_{ji} \leq 2$. The maximum value will be reached if distribution j is concentrated at the mean of distribution i.
(f) In general, the higher the overlapping index O_{ji} the lower will be O_{ij}. That is, the more group j is included in the range of distribution i, the less distribution j is expected to be included in the range of i.

Properties (a) to (f) show that O_{ji} is an index that measures the extent to which group j is included in the range of group i and the degree to which group i out-span group j.

The index O_i in equation (7.1) is a weighted average of the individual overlapping indexes. That is,

$$O_i = \sum_j p_j O_{ji} \, , \tag{7.5}$$

is an index indicating how the overlapping of the overall distribution by distribution i can be constructed.

It is clear that the indices of overlapping are not independent. Moreover, they are correlated with intra-group inequality indices. This is so because an increase in the variability of group i may increase the range, and therefore the probability of overlapping with other groups. Hence, it is worthwhile to present the contribution of each component to overall inequality.

7.2. Between-group variability

I will use the term GB to denote the between-group GMD as developed in Yitzhaki and Lerman (1991). It is defined as

$$GB = 2\,\mathrm{COV}(\mu_i, F_{oi})\,. \tag{7.6}$$

In words, GB is twice the covariance between the mean variate of each group and its mean rank in the overall population. This definition differs from the usual definition used in the literature of income inequality by Dagum (1997), Pyatt (1976), Shorrocks (1983a) and Silber (1989). See also Dagum (1987). In their definition, the between-group term is based on the covariance between the mean income of each group and the rank of the mean income of each group in the overall ranking. Here the covariance is between mean income and mean rank of the observations. It can be shown that the between-group GMD in this paper is never greater than the one defined in the above-mentioned papers.

The properties of the between-group GMD are the following:

(a) The between-group component can be zero either if mean variates are equal, or if mean ranks are equal. This definition is different from the definition of between-group variability that is embedded in the variance.
(b) The between-group component can, in some extreme cases, be negative. That is, the covariance between the average value of the groups and the average rank of the groups in the overall ranking can be negative. For example, one group may be composed of many small values of X and one extremely large observation, so that the mean of the variate is high but the mean rank is low.
(c) When groups are perfectly stratified, the upper bound for between-group variability is reached. The index in this case is identical to what is obtained by ranking groups according to their mean income, which is Pyatt's (1976) definition of between-group inequality. The difference between the upper bound and definition (7.6) can be used to evaluate the impact of overlapping on between-group inequality.

Since overall variability is given, the higher the stratification the higher the between-group inequality and the lower the overlapping. This property can be used to reflect the robustness of grouping. The ratio of the between-group component to its maximum value (under perfect stratification) can serve as an index of the robustness of the grouping (Milanovic and Yitzhaki, 2002).

Figures 3a and 3b demonstrate the difference between GMD and variance decomposition. Figure 3a presents two asymmetric distributions that are identical except for the direction of the asymmetry. (The overall distribution is $1 \cup 2$.) Figure 3b, presents the same distributions, except that each distribution is twisted 180° around its mean. The resulting overall distribution is $1' \cup 2'$. Since

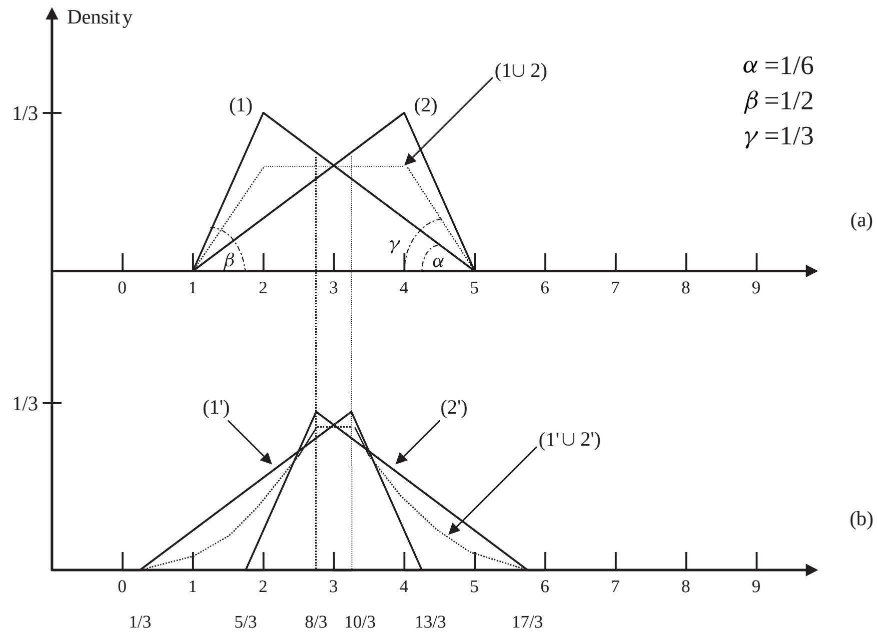

Figure 3. Densities with equal inter and intra-group variance; equal intra-group GMD.

the distributions are identical in shape, the intra-group variances and GMD's are identical. The means of each distribution and its twisted distribution in figures 3a and 3b are identical, so that between group and over-all variances in the two figures are identical. The overlapping of distributions in figure 3a is larger than the overlapping in figure 3b. Therefore, we should expect the between-group GMD in figure 3b to be higher than between-group GMD at figure 3a (the difference in average rank is higher in 3b, while the means are identical). Hence, the GMD decomposition indicates that the grouping in figure 3b is more distinguishable than the grouping in figure 3a.

The decomposition of the GMD into the contribution of sub-groups reveals more information on the underlying distributions than the decomposition of the variance. The GMD is the appropriate index whenever stratification has some implications for the interpretations of the results.

8. Moments versus extended Gini

As pointed out earlier, the major controversy between the Italian school and the Western statisticians was whether to approach the description of the distribution function through the differences between observations or, alternatively, to use the moment generating function. Once the GMD is presented as a covariance, how to describe it as a moment of a distribution is clear. Moreover one can present the variance and the covariance as a result of differencing two i.i.d. variables. Thus, the difference between the schools of thought dissolves.

The variance is a central moment, that is, a linear function of the density function and a power function of the variate. The natural extension is

$$\mu(r) = E\{[x - \mu(1)]^r\}\,. \tag{8.1}$$

The extension of the GMD to resemble a moment is accomplished by reversing the role of the variate and the probabilities. Since one needs a monotonic function to have an inverse function, it is convenient to rely on the cumulative distribution function.

$$\mu - G(\nu) = \int [1 - F_x(x)]^\nu dx = \mu + \nu\,\mathrm{COV}(x, [1 - F_x(x)]^{\nu-1})\,. \tag{8.2}$$

If $\nu \to 1$, then (8.2) converges, by definition, to μ so that variability is ignored. If $\nu = 2$, then (8.2) is equal to $\mu - 2G_X$. As $\nu \to \infty$ (8.2) converges to the lower bound of the distribution and as $\nu \to 0$, (8.2) converges to the upper bound. Equation (8.2) represents a family of variability measures that is referred to as the extended Gini family in Yitzhaki, 1983. By selecting ν, the investigator selects the distance function with which he works. Sections 3-7, can be generalized so that the G-correlations and decompositions can be performed on the extended Gini. Such an extension is beyond the scope of this paper. For a first step in this direction, see Schechtman and Yitzhaki (2003).

The application of a power function to the variate, as in (8.1), changes the unit and in many cases one has to use the inverse of the power function to return to the original units (e.g., variance and standard deviation). Since $F(\ )$ is unit free, all extended Gini indices have the same unit, so that there is no need to “undo” the impact of the power function. Also, all extended Gini variability indices are special cases of Yaari’s (1988) dual approach to utility theory.

9. Moments versus distance between distributions

Can one produce convergence theorems for distributions using the GMD? The answer to this question was given in the late twenties and the early thirties, although it seems to have been forgotten.

Let $F_n(X)$ be the empirical distribution function. Cramér (1928, p. 144-147) suggested the following statistic for the distance of the cumulative distribution from its expected value as a criterion for testing the goodness of fit of a distribution

$$\frac{G_x}{n} = E\left\{\int [F_n(x) - F(x)]^2 dx\right\} . \tag{9.1}$$

Cramer suggested the right hand-expression. He was not aware that it is equal to the GMD. See Darling (1957). For a proof that the term is equal to GMD, see Serfling (1980, p. 57), or Yitzhaki (1998, p. 17).

Equation (9.1) can be adjusted to describe the distance between two empirical distributions. Those extensions are beyond the scope of this paper.

Smirnov (1937) modified Equation (9.1) to

$$w^2 = n\int [F_n(x) - F(x)]^2 dF(x) . \tag{9.2}$$

Changing the integration from dx to $dF(x)$ changes the properties of the statistic from being asymptotically distribution-free to being distribution-free and, of course, (9.2) is not related to the GMD. It is worth noting that Equation (9.1) can be adjusted to describe the distance between two empirical distributions. Those extensions are beyond the scope of this paper.

10. Conclusions

Without a doubt the variance is more convenient to use as a measure of variability. Because it is a parameter of the normal distribution, there is no chance for an alternative measure of variability to outperform it in describing this class of distributions. However, there are many areas in economics and in other social sciences in which the normal distribution and even a symmetric density function do not provide a good approximation to the data. In those cases, it may be that other measures of variability can outperform the variance on a number of dimensions. For example, the covariance imposes a symmetric relationship between the marginal distributions. This property is convenient and simplifies the analysis but it is not required for economic analysis that is mainly dealing with asymmetric relationships among variables. The imposition of symmetric relationship among variables results in a loss of information, as can be seen from the decomposition of the GMD. In other words, the decomposition of the GMD, whether one is dealing with decomposing the GMD of a sum of random variables, or a union of several groups, yields additional terms that can help us in describing the properties of the underlying distributions. Other dimensions are the sensitivity to extreme observations, and the issue of stochastic dominance.

The GMD is similar to the variance. The use of the GMD in case of a normal distribution may result in an inefficient method of estimation. But the increase in the size of data sets implies that the relative disadvantage of this property has declined over time. This can be viewed as a price paid for using the GMD, while the price associated with the use of the variance is the loss of information.

Provided that one is ready to impose certain restrictions on the distributions involved (exchangeability or stratification) then one can achieve corresponding properties for the analysis that relies on the GMD and the analysis that relies on the variance. This means that by imposing those restrictions, one can replicate every model that relies on the variance with a model that is based on the GMD, (by replacing every variance by the square of the GMD, and every Pearson's correlation coefficient by the G-correlation, see Shalit and Yitzhaki (2002) for such an exercise in portfolio theory and Schechtman and Yitzhaki (2002) in multiple regression). Then, one can relax the restrictions that led to similarity between the GMD and the variance, to see whether there is an effect on the conclusion. In this sense the use of the GMD is superior to the use of the variance. The only price paid is a loss of efficiency in cases that are close to the normal distribution.

Further research is needed to be able to evaluate the contribution of the insights added by using the GMD. Non-friendly Computer programs that can decompose the GMD of a sum of random variables, and the Gini coefficient into population sub-groups, together with statistical tests will be supplied by the author upon request.

ACKNOWLEDGMENTS

Partial financing from the German-Israeli Fund is gratefully acknowledged. I am indebted to Reuben Gronau, Samuel Kotz, Peter Lambert, Ingram Olkin, Edna Schechtman and Joel Slemrod and participants of a conference at Bocconi University for many helpful comments. I am grateful to two anonymous referees who pointed out the contribution of the enormous Italian literature.

The literature on the GMD, has developed over more than a century, in different contexts, disciplines, languages, and journals. It is not surprising that many of the ideas were developed independently (see David (1968, 1981), Giorgi and Pallini (1987) and Harter (1978), Simpson (1948) among many others). Sometimes it is really hard to determine what is required in order to attribute an equation or a concept to a specific person or paper: is it writing an equation by passing, or do we require a full understanding of its implications. I apologize in case I failed to give the credit to the person or paper that deserves it.

REFERENCES

Barnett, V., Green, P. J., and Robinson, A. (1976) Concomitants and correlation estimates, *Biometrika*, 63, 2, 323-328.

Bassett, G. Jr. and Koenker, R. (1978) Asymptotic theory of least absolute error regression, *Journal of the American Statistical Association*, 73, 618-622.

Beach, C. M. and Davidson, R. (1983) Distribution-free statistical inference with Lorenz curves and income shares, *Review of Economic Studies*, 50, 723-735.

BISHOP, J. A., CHOW, K. V., and FORMBY, J. P. (1994) Testing for marginal changes in income distributions with Lorenz and concentration curves, *International Economic Review*, 35, 2, (May), 479-88.

BISHOP, J. A., FORMBY, J. P., and THISTLE, P. D. (1992) Convergence of the south and non-south income distributions, 1969-1979, *American Economic Review*, 82, 1, (March), 262-72.

BLITZ, R. C. and BRITTAIN, J. A. (1964) An extension of Lorenz diagram to the correlation of two variables, *Metron*, 137-143.

CRAMÉR, H. (1928) On the composition of elementary errors, second paper: statistical applications, *Skandinavisk Actuareitidskrift*, 11, 141-180.

DAGUM, C. (1987) Measuring the economic affluence between populations of income receivers, *Journal of Business & Economic Statistics*, 5, 1, 5-12.

DAGUM, C. (1997) Decomposition and interpretation of Gini and generalized entropy inequality measures, *Proceedings of the American Statistical Association*, Business and Economic Section, 200-205.

DARLING, D. A. (1957) The Kolmogorov-Smirnov, Cramer-von Mises tests, *Annals of Mathematical Statistics*, 28, 823-838.

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

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

DAVIDSON, R. and DUCLOS, J. Y. (1997) Statistical inference for the measurement of the incidence of taxes and transfers, *Econometrica*, 65, 6, 1453-1465.

DEATON, A. (1979) The Distance function in consumer behavior with applications to index numbers and optimal taxation, *The Review of Economic Studies*, 46, 3, (July), 391-405.

DE VEAUX (1976) *Technical Report Number 5*, SIAM Institute for Mathematics and Society, Stanford University.

FELDSTEIN, M. S. (1969) Mean variance analysis in the theory of liquidity preference and portfolio selection, *Review of Economic Studies*, 36, (January), 335-46.

FORMBY, J. P., SMITH, W. J., and ZHENG, B. (1999) The coefficient of variation, stochastic dominance and inequality: A new interpretation, *Economics Letters*, 62, 319-323.

FOSTER, J. and OK, E. A. (1999) Lorenz dominance and the variance of the logarithms, *Econometrica*, 67, 4, 901-907.

EUBANK, R., SCHECHTMAN, E., and YITZHAKI, S. (1993) A test for second order dominance, *Communications in Statistics, Theory and Methods*, 61, 7, 1893-1905.

GASTWIRTH, J. L. (1971) A general definition of the lorenz curve, *Econometrica*, 39, 1037-39.

GASTWIRTH, J. (1972) The estimation of the Lorenz curve and Gini index, *The Review of Economics and Statistics*, 54, 306-316.

GINI, C. (1912) *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*, in Studi Economico-giuridici della Regia Facoltà Giurisprudenza, anno III, parte II, Cuppini, Bologna.

GINI, C. (1965) On the characteristics of Italian statistics, *Journal of the Royal Statistical Society*, Series A, 128, 1, 89-109.

GINI, C. (1966) *Statistical Methods*, Instituto Di Statistica E Ricerca Sociale "Corrado Gini", Universita Degli Studi di Roma, Roma.

GIORGI, G. M. (1990) Bibliographic portrait of the Gini concentration ratio, *Metron*, 48, 1-4, 183-221.

GIORGI, G. M. and PALLINI, A. (1987) About a general method for the lower and upper distribution-free bounds on Gini's concentration ratio from grouped data, *Statistica*, 47, 2, 171-184.

GOLDIE, C. M. (1977) Convergence theorems for empirical lorenz curves and their inverses, *Advances in Applied Probability*, 9, 765-791.

GRETHER, D. M. (1974) Correlations with ordinal data, *Journal of Econometrics*, 2, 241-246.

HADAR, J. and RUSSELL, W. R. (1969) Rules for ordering uncertain prospects, *American Economic Review*, 59, 25-34.

HANOCH, G. and LEVY, H. (1969) The efficiency analysis of choices involving risk, *Review of Economic Studies*, 36, 335-346.

HART, P. E. (1975) Moment distributions in economics: an exposition, *Journal of The Royal Statistical Society*, Series A, 138, 3, 423-434.

HARTER, H. L. (1978) *A chronological annotated bibliography of order statistics*, vol. 1, Pre-1950, Government Printing Office, Washington, DC: U. S.

HOEFFDING, W. (1948) A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics*, 19, 293-325.

JORGENSON, D. W. and SLESNICK, D. T. (1984) Aggregate consumer behaviour and measurement of inequality, *Review of Economic Studies*, 51, 3, (July), 369-392.

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

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

KENDALL, M. G. (1955) *Rank correlation methods*, (second edition), Hafner Publishing Company, New York.

KOTZ, S. and SEEGER, J. P. (1991) *A New approach to dependence in multivariate distributions*, in Advances in Probability Distributions with Given Marginals, Kluwer Academic Publishers.

LAMBERT, P. (1993) *The distribution and redistribution of income*, (second edition), Manchester University Press, Manchester.

LASSWELL, T. E. (1965) *Class and stratum*, Houghton Mifflin, Boston.

LEVY, H. and MARKOWITZ, H. M. (1979) Approximating expected utility by a function of mean and variance, *American Economic Review*, 78, (June), 308-317.

LEVY, H. (1992) Stochastic dominance and expected utility: survey and analysis, *Management Science*, 38, 4, 555-593.

LORENZ, M. O. (1905) Methods for measuring concentration of wealth, *Journal of the American Statistical Association*, New Series 70, 209-219.

MCKENZIE, G. W. and PEARCE, I. F. (1982) Welfare measurement: a synthesis, *American Economic Review*, 72, 4, (September), 669-682.

MILANOVIC, B. and YITZHAKI, S. (2002) Decomposing World Income Distribution: Does the World Have a Middle Class, *Review of Income and Wealth*, 48, 2, (June), 155-178.

NELSEN, R. B. (1999) *An introduction to copulas*, Springer-Verlag, New York.

NYGÂRD, F. and SANDSTRÖM A. (1981) *Measuring income inequality*, Almqvist & Wiksell International, Stockholm.

OLKIN, I. and YITZHAKI, S. (1992) Gini regression analysis, *International Statistical Review*, 60, 2, (August), 185-196.

PIETRA, G. (1915) Delle relazioni tra gli indici di variabilita: I, II, *Atti del reale Instituto Veneto di Scienze, Lettere ed Arti, a.a. 1914-15*, 74, 2, 775-804.

PIETRA, G. (1931) Contributions à la méthodologie du calcul du rapport de concentration, *Bulletin de l'Institut International de Statistique*, XX session, (Madrid, 1931), tome 26, 2ème Livrasion, 568-574.

PYATT, G. (1976) On the interpretation and disaggregation of Gini coefficients, *The Economic Journal*, 86, 243-254.

RAMASUBBAN, T. A. (1958) The mean differences and the mean deviations of some discontinuous distributions, *Biometrika*, 45, 56-549.

RAMASUBBAN, T. A. (1959) The generalized mean differences of the binominal and poisson distributions, *Biometrika*, 46, 29-223.

RAMASUBBAN, T. A. (1960) Some distributions arising in the study of generalized mean differences, *Biometrika*, 47, 73-469.

RAO, V. M. (1969) Two decompositions of concentration ratio, *Journal of the Royal Statistical Society*, Series A, 132, 418-425.

ROCKINGER, M. and JONDEAU (2001) *Conditional dependency of financial series: An applications of copulas*, Working Paper.

ROTHSCHILD, M. and STIGLITZ, J. E. (1970) Increasing risk: I: a definition, *Journal of Economic Theory*, 2, 225-353.

RUNCIMAN, W. G. (1966) *Relative deprivation and social justice*, Routledge and Kegan Paul, London.

SHALIT, H. and YITZHAKI, S. (1984) Mean-Gini, portfolio theory, and the pricing of risky assets, *Journal of Finance*, 39, 5, (December), 1449-68.

SHALIT, H. and YITZHAKI, S. (2002) *Derivation of the mean-Gini efficient portfolio frontier*, Discussion Paper No. 02-17, Ben-Gurion University, October.

SCHECHTMAN, E. (2000) *Stratification: measuring and inference*, Mimeo.

SCHECHTMAN, E. and YITZHAKI, S. (1987) A Measure of association based on Gini's mean difference, *Communications in Statistics Theory and Methods*, Series A, 16, 1, 207-231.

SCHECHTMAN, E. and YITZHAKI, S. (1999) On the proper bounds of the Gini correlation, *Economics Letters*, 63, 2, (May), 133-138.

SCHECHTMAN, E. and YITZHAKI, S. (2000) *Gini's multiple regressions*, Mimeo.

SCHECHTMAN, E. and YITZHAKI, S. (2003) A family of correlation coefficients based on extended Gini, *Journal of Economic Inequality*, forthcoming.

SERFLING, R. J. (1980) *Approximation theorems of mathematical statistics*, John Wiley and Sons, New York.

SHIH, W. J. and HUANG, W. M. (1992) Evaluating correlation with proper bounds, *Biometrics*, 48, 1207-1213.

SHORROCKS, A. F. (1983a) The impact of income components on the distribution of family incomes, *Quarterly Journal of Economics*, 98, 311-326.

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

SILBER, J. (1989) Factor components, population subgroups, and the computation of the Gini index of inequality, *Review of Economics and Statistics*, 71, 107-115.

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

SMIRNOV, N. V. (1937) On the distribution of the w^2 criterion of von mises, *Rec. Math.*, 2, 973-993.

STUART, A. (1954) The correlation between variate values and ranks in samples from a continuous distribution, *British Journal of Statistical Psychology*, 7, 37-44.

STUART, A. and ORD, J. K. (1987) *Kendall's advanced theory of statistics*, 1, (fifth edition), Charles Griffin & Company, London.

SUITS, D. B. (1977) Measurement of tax progressivity, *American Economic Review*, 67, 747-752.

TOBIN, J. (1958) Liquidity preference as behavior toward risk, *Review of Economic Studies*, 25, (February), 65-87.

TOBIN, J. (1965) The theory of portfolio selection, in Frank H. Hahn and Frank P. R., Brechling eds., *Theory of Interest Rates*, MacMillan, London.

YAARI, M. E. (1988) A controversial proposal concerning inequality measuremen, *Journal of Economic Theory*, 44, 2, 381-397.

YITZHAKI, S. (1982) Stochastic dominance, mean variance and Gini's mean difference, *American Economic Review*, 72, 178-185.

YITZHAKI, S. (1983) On an extension of the Gini inequality index, *International Economic Review*, 24, 617-618.

YITZHAKI, S. (1990) On the sensitivity of a regression coefficient to monotonic transformations, *Econometric Theory*, 6, 2, 165-169.

YITZHAKI, S. (1994) Economic distance and overlapping of distributions, *Journal of Econometrics*, 61, 147-159.

YITZHAKI, S. (1996) On using linear regression in welfare economics, *Journal of Business & Economic Statistics*, 14, 4, (October), 478-86.

YITZHAKI, S. (1998) More than a dozen alternative ways of spelling Gini, in D. J. Slottje, ed., *Research on economic inequality*, 8, 13-30.

YITZHAKI, S. (1999) Necessary and sufficient conditions for dominance using generalized Lorenz curves, in D. J. Slottje, (ed.), *Advances in econometrics, income distribution and scientific methodology: Essays in honor of Camilo Dagum*, Physica-Verlag, Heidelberg.

YITZHAKI, S. and EISENSTAEDT, M. (2003) Groups' Ranking versus Individuals' Ranking, in: *Fiscal policy, inequality, and welfare, Research on economic inequality*, 10, (eds. Y. Amiel and J. H. Bishop), 10, JAI, Amsterdam, 101-123.

YITZHAKI, S. and LERMAN, R. (1991) Income stratification and income inequality, *Review of Income and Wealth*, 37, 313-329.

YITZHAKI, S. and OLKIN, I. (1988) *Concentration curves*, Working Paper No. 179, Dept. of Economics, The Hebrew University.

YITZHAKI, S. and OLKIN, I. (1991) Concentration indices and concentration curves, in K. Mosler and M. Scarsini (eds.), *Stochastic Orders and Decisions under Risk, Institute of Mathematical Statistics: Lecture-Notes Monograph Series, 19*, 1991, 380-392.

YITZHAKI, S. and SLEMROD, J. (1991) Welfare dominance: An application to commodity ctxation, *American Economic Review*, 81, 3, (June), 480-496.

YITZHAKI, S. and WODON, Q. (2000) *Mobility, inequality, and horizontal inequity*, The World Bank, mimeo.

SHLOMO YITZHAKI
Department of Economics
Hebrew University
Jerusalem, 91905 (Israel)
msruhama@mscc.huji.ac.il
and
Central Bureau of Statistics
Jerusalem (Israel)