University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers

Charles M. Beach

MODEL-FREE STATISTICAL INFERENCE WITH LORENZ CURVES, INCOME SHARES, AND GINI COEFFICIENTS

DP #577-79

Model-Free Statistical Inference with Lorenz Curves,

Income Shares, and Gini Coefficients

Charles M. Beach

Queen's University

September 1979

The author wishes to thank A. B. Atkinson, F. A. Cowell, A. S. Goldberger, and A. F. Shorrocks for helpful discussion; he himself retains full responsibility for all shortcomings of the paper. The paper was written during visits to the London School of Economics and Political Science and the Institute for Research on Poverty, University of Wisconsin - Madison. Partial support was provided by the Social Sciences and Humanities Research Council of Canada.

Abstract

This paper (i) derives asymptotic distributions for income share statistics, Lorenz curve ordinates, and estimated Gini coefficients, and (ii) thence develops simple statistical inference procedures for these commonly used tools of applied distribution analysis.

The basic concepts underlying this work are population income quantiles ξ_{p_i} defined by $p_i = \int_0^{\xi_{p_i}} f(y)dy$ where $f(y)$ is the population income density and $0 < p_i < 1$. These are estimated by sample quantiles $\hat{\xi}_{p_i}$ based on a sample of size N. Under general conditions, a vector of K sample quantiles, $\hat{\xi}$, is asymptotically normal with mean ξ and covariance matrix $N^{-1}\Sigma$ where $\Sigma_{ij} = p_i(1-p_j)/f(\xi_{p_i})f(\xi_{p_j})$ for $p_i < p_j$. Since a vector of K sample Lorenz curve ordinates $\hat{\Phi}$ can be expressed as differentiable functions of $\hat{\xi}$, the asymptotic distribution of $\hat{\Phi}$ is also normal with mean Φ and covariance matrix N^{-1} V where $V_{ij} = \xi_{p_i} \xi_{p_j} p_i (1-p_j)/\mu^2$ which does not depend on $f(\xi_{p_i})$, so that distribution- or model-free inferences can be carried out.

The paper then extends this result to income share statistics and Gini coefficients, and then illustrates a number of computationally simple statistical tests based on these results.

I. Introduction

One of the most frequently used devices to describe and compare distributional inequality in economics is the Lorenz curve. It has intuitive appeal and can be easily estimated. It is generally defined and not dependent on any prior specification of an underlying distribution function. It is the basis of a necessary and sufficient condition for ranking two distributions independent of specific utility functions [2]. It is also the basis for several summary measures of income (or wealth) inequality such as the Gini concentration coefficient, perhaps the most frequently used single measure of inequality. Finally, the Lorenz curve also provides a disaggregated overview of the share structure of inequality in a distribution, so that one can see over which regions of a distribution inequality is relatively marked.

So far, however, Lorenz curves and income shares have been used essentially as descriptive devices and not as tools for rigorous statistical inference. This is at least in part due to the complexities of the sampling distributions associated with these devices, but is also partly due a surprising lack of inquiry into the problem of formalizing statistical inference with Lorenz curves. Such a state of affairs is particularly troublesome in light of the massive outflow of recent empirical work using micro data to compare income and wealth inequality in different distributions, and of the current general interest in distributional considerations. This paper offers a solution to this problem by forwarding a new approach to distributional inference based on quantile analysis and the asymptotic distribution of sample income quantiles. Indeed, it will be shown that statistical inferences with Lorenz curves, income shares, and Gini coefficients are (asymptotically) distribution-free or model-free in the sense that they do not require knowledge of the underlying distribution model or parent distribution of the sample.

So far, statistical inference and confidence intervals have been worked out only for a few summary inequality measures [16, 18]. But such measures frequently hide much interesting distributional detail, and contain implicit value norms that may not be adequately recognized or generally acceptable. The present paper is written in the spirit of these studies, but extends the analysis to disaggregated inequality levels so as to permit a much richer and more detailed understanding of the structure of inequality in a distribution. As a useful corollary, the analysis also provides for inferences and standard errors of the Gini coefficient as well.

This paper focuses on the problem of disaggregated statistical inference, and for convenience and clarity we will assume that we are working with samples of micro data. The approach thus contrasts with that of Gastwirth [15] and Gastwirth and Glauberman [17], who focus on interpolation methods for estimation of Lorenz curves and thus on "interpolation error" as opposed to "sampling error". In contrast to Gastwirth [16] and Kakwani [18], the present approach is disaggregative in orientation and leads to model-free inferences -- unlike maximum likelihood procedures, for example. And, in contrast to Kakwani and Podder [19, 20] and Thurow [32], the current approach does not require any curve-fitting or iterative nonlinear estimation techniques in order to carry out inferences on Lorenz curves and income shares. The approach also avoids the need to fit specific distribution models or density functions to empirical distributions in order to extract the relevant inequality information from the data -- again in contrast to analyses, for example, by Aigner and Goldberger [1] and Kloek and van Dijk [22, 23]. The present work, however, can be seen as an extension of the model-free approach of Beach [4] of basing distributional analysis on a set of income quantiles, so that the overall structure of inequality in a distribution can be studied without the need of fitting specific functional forms.

The objectives of the paper are thus (i) to draw economists' attention to a body of statistical theory on sample quantiles that can be usefully exploited in distributional analysis; and (ii) to provide model-free inference techniques for Lorenz curves, income shares, and Gini coefficients.

The paper proceeds as follows. The next section introduces income quantiles and reviews some of the basic sampling theory to be used. Sections III and IV apply the theory to derive asymptotic distributions of Lorenz curve ordinates, income shares, and Gini coefficients. Sections V and VI then illustrate various inference procedures, and a few general comments are provided in the brief concluding section.

II. Review of Sampling Distributions of Income Quantiles

II.1) Lorenz Curves and Quantiles

In order to define a Lorenz curve conveniently, let $f(y)$ be the (continuous) parent density function of income recipients. Then the proportion of recipients with incomes up to y is the (cumulative) distribution function (or c.d.f.)

$$F(y) = \int_{-\infty}^{y} f(u)\, du \qquad (2.1)$$

and the proportion of total income receipts in the distribution by recipients with incomes up to y is the incomplete (first) moment function

$$\Phi(y) = \frac{1}{\mu} \int_{-\infty}^{y} u\, f(u) du \qquad (2.2)$$

where the mean income level, μ, is assumed to exist. Then just as the Lorenz curve abscissa $F(y)$ varies from 0 to 1, the Lorenz curve ordinate $\Phi(y)$ also varies from 0 to 1 monotonically where we assume, for convenience, that all incomes are positive. The so-called curve of concentration or Lorenz curve is the function $\Phi(F)$ defined parametrically in terms of income levels y by (2.1) and (2.2).[1]

An income quantile ξ_p corresponding to abscissa value $p(0 \leq p \leq 1)$ on a Lorenz curve is defined implicitly by $p = \int_0^{\xi_p} f(u)du$ or $F(\xi_p) = p$ where $F(y)$ is assumed to be strictly monotonic. For example, the first decile level is $\xi_{.1}$ such that $.1 = \int_0^{\xi_{.1}} f(u)du$, and the median income level is $\xi_{.5}$ such that $.5 = \int_0^{\xi_{.5}} f(u)du$, so that half the receipients have incomes less than or equal to $\xi_{.5}$ and half have more.[2] Thus, corresponding to a set of K abscissas $p_1 < p_2 < \ldots < p_K$, we have a set of K population income quantiles

---

1. For an explicit definition of Φ in terms of F, see Gastwirth [14] and Dorfman [11].

2. It may be of interest to remark that concern with income quantiles has also recently developed in the theoretical literature on measuring economic inequality [31, p. 31; 10].

$\xi_{p_1} < \xi_{p_2} < \ldots < \xi_{p_K}$. Note that the ξ_{p_i}'s are not in general parameters of a distribution, but simply distribution characteristics which we seek to estimate by sample statistics. Consequently, while quantile procedures are "nonparametric", they are not necessarily "distribution-free" [7, p. 15]. Note also that the quantile abscissas, p_i, need not necessarily be equally spaced. We shall assume for convenience in this paper that they are (e.g., that the ξ_{p_i}'s are all deciles, centiles, or quartiles, say). But if one were particularly interested in upper and lower shares, for example, one might choose closer quantiles over those regions than elsewhere in the distribution.

II.2) Exact Distributions of Order Statistics

Consider a random sample of N observations drawn from the probability density model f(y) with corresponding c.d.f. F(y), and order the observations from the smallest to the largest. Then Y_ℓ in the ordered sample represents the ℓ'th smallest observation where $1 \leq \ell \leq N$. The probability that $(\ell-1)$ of the sample observations fall below a value y_ℓ, one falls in the range $y_\ell \pm \frac{1}{2}\ dy_\ell$, and the remaining $(N-\ell)$ fall above y_ℓ is then given [21, pp. 236, 252; 33, p. 236] by the probability element

$$dG(y_\ell) = \frac{N!}{(N-\ell)!(\ell-1)!}\,[F(y_\ell)]^{\ell-1}[1-F(y_\ell)]^{N-\ell}f(y_\ell)\ dy_\ell. \quad (2.3)$$

The corresponding mean and variance of the ℓ'th order-statistic, Y_ℓ, are thus given [28, p. 13] as

$$E(Y_\ell) = \frac{N!}{(N-\ell)!(\ell-1)!}\int_0^\infty u[F(u)]^{\ell-1}\,[1-F(u)]^{N-\ell}f(u)du$$

and

$$V(Y_\ell) = E(Y_\ell^2) - E(Y_\ell)^2$$

$$= \frac{N!}{(N-\ell)!(\ell-1)!} [\int_0^\infty u^2 [F(u)]^{\ell-1} [1-F(u)]^{N-\ell} f(u)\, du$$

$$- \{\int_0^\infty u[F(u)]^{\ell-1} [1-F(u)]^{N-\ell} f(u)du\}^2].$$

From these expressions it can be readily seen that exact sampling distributions for order-statistics have two important characteristics. First, the observations in an <u>ordered</u> sample will no longer be independent[3] or identically distributed even when the original sample observations were. Second, the exact sampling distributions of order statistics are relatively complicated to handle analytically and depend very directly upon the underlying parent density model $f(y)$, so that exact inferences about the parent quantiles ξ_{p_i} based on such order-statistics are not distribution-free or "model-free".[4]

---

3. Corresponding joint distributions and covariances for any two order statistics Y_ℓ and Y_k can also be found in [28, p. 13], [33, p. 236], and [21, pp. 270, 325].

4. It is worth noting, however, that pairs of order-statistics can be used to set distribution-free <u>confidence</u> <u>intervals</u> for population quantiles. In particular, it can be easily shown that, if $F(y_\ell) \leq p \leq F(y_k)$,

$$\text{Prob}(Y_\ell \leq \xi_p \leq Y_k) = \Sigma_{j=\ell}^{k-1} \binom{N}{j} p^j (1-p)^{N-j}$$

for order-statistics Y_ℓ, Y_k [33, pp. 330-331; 21, pp. 517-18]. However, as we shall want to work with functions or transforms of sample quantiles and obtain smooth confidence bands for the set of transformed quantiles, we shall deal directly with their sampling distribution functions and not just with confidence intervals for conveniently selected order-statistics.

II.3 Asymptotic Distribution of Sample Quantiles

An asymptotic approximation to the distribution of sample quantiles, however, does provide the basis for distribution-free inference for sample shares and Lorenz curve ordinates. Given a random sample of N observations,[5] define an estimate of the population ξ_p to be

$$\begin{aligned} \hat{\xi}_p &= Y_{Np} \text{ if } Np \text{ is an integer} \\ &= Y_{[Np]+1} \text{ if } Np \text{ is not an integer,} \end{aligned} \tag{2.4}$$

where [Np] denotes the greatest integer not exceeding Np. These corresponding sample quantiles are known to have several useful statistical properties.

In particular, it can be shown that, if F(y) is strictly monotonic, $\hat{\xi}_p$ defined in (2.4) has the property of strong or almost sure consistency [27, p. 355]; that is, $\lim_{N\to\infty} \hat{\xi}_p = \xi_p$ with probability one, so that a fortiori it is weakly consistent as well. In addition, the $\hat{\xi}_{p_i}$'s are also asymptotically normal with a relatively simple covariance structure. More formally, we state this result (without proof) as the basic cornerstone of this paper.

---

5. Since this paper is concerned essentially with statistical inference and not estimation, it is assumed throughout that the analyst has access to actual micro data. If, however, he does not and the distribution data are available only in interval or histogram form, then standard interpolation procedures must be employed to obtain estimates of quantile income levels and income shares (e.g., Gastwirth [15]). In this case, interpolation errors occur in addition to sampling errors in estimating the ξ_{p_i} and in computing asymptotic standard errors.

<u>Theorem 1</u>:

Suppose that, for the set of proportions $\{p_i\}$ such that $0 < p_1 < p_2 < \ldots < p_K < 1$, $\hat{\xi} = (\hat{\xi}_{p_1}, \hat{\xi}_{p_2}, \ldots, \hat{\xi}_{p_K})'$ is a vector of K sample quantiles from a random sample of size N drawn from a continuous population density f(y) such that the ξ_{p_i}'s are uniquely defined and $f_i \equiv f(\xi_{p_i}) > 0$ for all $i = 1, \ldots, K$. Then the vector $\sqrt{N}\,(\hat{\xi}-\xi)$ converges in distribution to a K-variate normal distribution with mean zero and covariance matrix Λ. That is, $\hat{\xi}$ is asymptotically normal with mean vector $\xi = (\xi_{p_1}, \xi_{p_2}, \ldots, \xi_{p_K})'$ and asymptotic covariance matrix $(1/N)\,\Lambda$ where

$$\Lambda = \begin{bmatrix} \frac{p_1(1-p_1)}{f_1^2} & \cdots & \frac{p_1(1-p_K)}{f_1 f_K} \\ \vdots & & \vdots \\ \frac{p_1(1-p_K)}{f_1 f_K} & \cdots & \frac{p_K(1-p_K)}{f_K^2} \end{bmatrix} . \qquad (2.5a)$$

If P denotes the matrix

$$p = \begin{bmatrix} p_1 & p_1 & \cdots & p_1 \\ p_1 & p_2 & \cdots & p_2 \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \cdots & p_K \end{bmatrix} ,$$

$p' = (p_1, \ldots, p_K)$, and $F = \mathrm{Diag}\,[f_1, \ldots, f_K]$, Λ can be expressed in matrix form as

$$\Lambda = F^{-1}\,[P - pp']F^{-1} \qquad (2.5b)$$

Proofs of Theorem 1 can be found, for example, in Wilks [33, pp. 273-74], and Kendall and Stuart [21 , pp. 237-39].[6] Since $\hat{\xi}_{p_i}$ is a consistent estimate of ξ_{p_i}, one can of course calculate a consistent asymptotic standard error of $\hat{\xi}_{p_i}$ as $[p_i(1-p_i)/Nf(\hat{\xi}_{p_i})^2]^{\frac{1}{2}}$.

It is important to note, however, that asymptotic inference on quantile income levels still requires knowledge of the underlying distribution model $f(\cdot)$ in computing the standard errors. It is thus desirable to work with transforms of these quantile variables which will allow model-free inferences. We now make use of Theorem 1 in deriving asymptotic distributions of sample share statistics and Lorenz curve ordinates.

III. Income Shares and Lorenz Curves

III.1) Asymptotic Distribution of Lorenz Curve Ordinates

To estimate Lorenz curve ordinates, recall first of all from (2.2) that

$$\Phi(\xi_{p_i}) = \frac{1}{\mu} \int_0^{\xi_{p_i}} uf(u)du = \frac{F(\xi_{p_i})}{\mu} \int_0^{\xi_{p_i}} \frac{uf(u)du}{F(\xi_{p_i})}$$

$$= p_i \cdot \frac{E(Y|Y \leq \xi_{p_i})}{E(Y)} = \frac{\tau_i}{\mu} .$$

Consequently, the sample estimate of $\Phi(\xi_i)$ may be computed as

$$\hat{\Phi}_i = \Sigma_{Y_j \leq \hat{\xi}_{p_i}} Y_j / \Sigma_{j=1}^N Y_j \doteq p_i \left(\frac{\bar{Y}_{\hat{\xi}_{p_i}}}{\bar{Y}}\right), \; i = 1, \ldots, K \qquad (3.1)$$

where $\bar{Y}_{\hat{\xi}_{p_i}} = \Sigma_{Y_j \leq \hat{\xi}_{p_i}} Y_j/n_i$ and $n_i = [Np_i]$. This will be referred to as the feasible or sample estimator of $\Phi(\xi_{p_i})$.

---

6. Stronger and broader results than Theorem 1 can also be found in [8, pp. 56,58] and [3].

It will also be convenient to define the population income share function evaluated at the sample quantile estimate as

$$\Phi(\hat{\xi}_{p_i}) = \frac{1}{\mu} \int_0^{\hat{\xi}_{p_i}} uf(u)du. \tag{3.2}$$

While this is a random variable, since it depends upon $\hat{\xi}_{p_i}$, it is also clearly dependent on the (unknown) population distribution function. This will be referred to as the <u>infeasible</u> estimator of $\Phi(\xi_{p_i})$. A Lorenz curve in this paper is represented by a set of K ordinates $\{\Phi(\xi_{p_i})\}$ which are to be estimated from the sample. The line of argument of this section involves, first, establishing the asymptotic distribution of the infeasible estimators $\Phi(\hat{\xi}_{p_i})$ for i = 1, ..., K as transforms of the sample quantiles (Lemma 1); then arguing that $\hat{\Phi}_i$ and $\Phi(\hat{\xi}_{p_i})$ have the same limiting distribution (Lemma 2); and thence concluding that the asymptotic distribution of the feasible estimators $\hat{\Phi}_i$, i=1, ..., K, is exactly that derived for the $\Phi(\hat{\xi}_{p_i})$'s.

In order to derive the asymptotic distribution of a set of Lorenz curve ordinates $\{\Phi(\hat{\xi}_{p_i})\}$, it is useful first of all to recall the following result [27, p. 321] on limiting distribution of continuous functions of random variables. Suppose that T_N is a K-dimensional statistic $(t_{1N}, t_{2N}, \ldots, t_{KN})'$ and $\theta = (\theta_1, \ldots, \theta_K)'$ a corresponding vector of constants such that the limiting distribution of the scaled vector $\sqrt{N}(T_N - \theta)$ is a K-variate normal with mean zero and covariance matrix Σ. Suppose also that a scalar function of the statistic vector T_n, $g(T_N)$, is totally differentiable. Then it follows that the limiting distribution of $\sqrt{N}(g(T_N) - g(\theta))$ is also normal with mean zero and variance $v = j' \Sigma j$ where

$$j = [\frac{\partial g(T_N)}{\partial t_{1N}}, \ldots, \frac{\partial g(T_N)}{\partial t_{KN}}]' \Bigg|_\theta$$

is the gradient vector of $g(\cdot)$ evaluated at θ. More generally, if $g = (g_1(T_N), \ldots, g_M(T_N))'$ is an M-dimensional vector-valued function with each g_i a function of the statistic vector T_N, and if each g_i is again totally differentiable, the M-dimensional vector $\sqrt{N}(g(T_N) - g(\theta))$ has an M-variate normal limiting distribution with zero mean and (M x M) covariance matrix $V = J \Sigma J'$, where

$$J = [J_{ij}] = [\frac{\partial g_i(T_N)}{\partial t_{jN}}] \Bigg|_\theta$$

is now an (M x K) matrix in which the i'th row contains the gradient of g_i, again evaluated at θ.

In order to apply these results to the present situation, let g_i, i = 1, ..., K, be the incomplete (first) moment function $\Phi(y)$ defined in (2.2). The gradient of the function (2.2) evaluated at the population value ξ_{p_i} can be seen to be simply $\xi_{p_i} f(\xi_{p_i})/\mu = (1/\mu)\xi_{p_i} f_i$. Consequently, setting $T_N = (\hat{\xi}_{p_1}, \ldots, \hat{\xi}_{p_K})'$, $\theta = (\xi_{p_1}, \ldots, \xi_{p_K})'$, $g(T_N) = (\Phi(\hat{\xi}_{p_1}), \ldots, \Phi(\hat{\xi}_{p_K}))'$, and $\Sigma = \Lambda$, we note that $J_L = \text{Diag}\,[(1/\mu)\xi_{p_1} f_1, \ldots, (1/\mu)\xi_{p_K} f_K]$, so that the variance of the limiting distribution corresponding to V in the case of Lorenz curve ordinates is

$$V_L = \begin{bmatrix} (\frac{\xi_{p_1}}{\mu})^2 p_1(1-p_1) & \cdot & \cdot & \cdot & (\frac{\xi_{p_1}\xi_{p_K}}{\mu^2})p_1(1-p_K) \\ \vdots & & & & \vdots \\ (\frac{\xi_{p_1}\xi_{p_K}}{\mu^2})p_1(1-p_K) & \cdot & \cdot & \cdot & (\frac{\xi_{p_K}}{\mu})^2 p_K(1-p_K) \end{bmatrix} \qquad (3.3a)$$

$$= R[P - pp']R \qquad (3.3b)$$

where $R = \mathrm{Diag}\,[\xi_{p_1}/\mu, \ldots, \xi_{p_K}/\mu]$. We thus have the result

Lemma 1: Under the conditions of Theorem 1, the (scaled) vector of infeasible Lorenz curve ordinate estimates with elements $\sqrt{N}(\Phi(\hat{\xi}_{p_i}) - \Phi_i)$ calculated from (3.2) is asymptotically K-variate normal with mean zero and covariance matrix V_L given in (3.3) Consequently, the (infeasible) Lorenz curve ordinates $\Phi(\hat{\xi}_{p_i})$ are asymptotically joint normal with mean $\Phi_i = \Phi(\xi_{p_i})$ and asymptotic covariance matrix $(1/N)V_L$.

So far, however, we have established the asymptotic distribution only of an infeasible set of estimators $\{\Phi(\hat{\xi}_{p_i})\}$ of the Lorenz curve ordinates. What are calculated from the sample are the feasible or sample estimates $\{\hat{\Phi}_i\}$ defined in (3.1). However, analogous to the results for Aitken generalized-least-squares estimators in econometrics, the feasible and infeasible estimators can be shown to be asymptotically equivalently distributed.

Lemma 2: Under the conditions of Theorem 1, if the population density has finite mean and variance, $\sqrt{N}(\hat{\Phi}_i - \Phi_i)$ and $\sqrt{N}(\Phi(\hat{\xi}_{p_i}) - \Phi_i)$ have the same limiting distributions. Proof of this result is based on a modification of Theorem 1 in Gastwirth [16] and is provided in the Appendix. Basically, the argument involves showing that the conditional and unconditional means, $\bar{Y}_{\hat{\xi}_{p_i}}$ and $\bar{Y}$, in (3.1) are both asymptotically normal with appropriate means and variances inspite of the fact that $\bar{Y}_{\hat{\xi}_{p_i}}$ is stochastically conditioned.

Combining Lemmas 1 and 2, one now has the principal result of this paper.

Theorem 2: Under the conditions of Lemma 2, the vector of sample estimators $\hat{\Phi} = (\hat{\Phi}_1, \ldots, \hat{\Phi}_K)$ of Lorenz curve ordinates is asymptotically normal in that

$\sqrt{N}(\hat{\Phi} - \Phi)$ has a limiting K-variate normal distribution with mean zero and covariance matrix V_L specified in (3.3).

Consequently, asymptotic standard errors for the sample estimates $\hat{\Phi}_i$ are given by

$$\sqrt{\frac{v_{ii}^L}{N}} = \left(\frac{\hat{\xi}_{p_i}}{\hat{\mu}}\right)\sqrt{\frac{p_i(1-p_i)}{N}} \qquad \text{for } i = 1, \ldots, K. \tag{3.4}$$

The important thing to note about V_L, of course, is that, in contrast to Λ, it does not require knowledge of the underlying model density function $f(\cdot)$. It depends solely upon the chosen proportions p_i, the population mean μ, and the population quantile income levels ξ_{p_i}, which can be estimated consistently from the sample. Thus statistical inferences about the Lorenz curve ordinates can be carried out without having to know or estimate the underlying model or parent density function. It is in this sense that we say that Lorenz curve inferences are model-free. It is perhaps interesting to remark that this distribution-free aspect of Lorenz curve inference in the statistical field usefully complements Atkinson's [2] Lorenz curve criterion in the field of welfare economics for making distributional inferences independent of the exact form of underlying utility functions as well. Consequently, one has further reason to be interested in using Lorenz curve analysis in applied distribution work.

It is worth noting that the present result implies that it is unnecessary for Lorenz curve inference to fit functional forms to empirical Lorenz curves as suggested, for example, by Kakwani and Podder [19, 20] and Thurow [32]. It also implies that, to make Lorenz curve inferences, it is unnecessary as well to fit various density functions to empirical distributions such as done in Aigner and Goldberger [1] and in Kloek and van Dijk [22, 23]. In addition, it suggests that, along with (cumulative) income shares and means, it is useful in applied work and published data also be provide estimates of income quantiles. Indeed, the only new information that will be required to compute standard errors and various test statistics for Lorenz curves is a set of income quantiles, $\{\hat{\xi}_{p_i}\}$.

Furthermore, note from (2.1) and (2.2) that the derivatives of the population Lorenz curve,

$$\frac{d\Phi}{dF} = \frac{d\Phi(y)/dy}{dF/dy} = \frac{(y/\mu)f(y)}{f(y)}$$

$$= y/\mu,$$

is the so-called relative-mean-income curve [21, p. 49; 24], which has a number of useful inequality properties in its own right. Corresponding to the abscissa points $p_1, p_2, \ldots, p_K$, the relative-mean-income curve ordinates are thus $\xi_{p_1}/\mu, \xi_{p_2}/\mu, \ldots, \xi_{p_K}/\mu$.[7] It can be seen, then, that the elements of covariance matrix (3.3) are simply the products of selected proportions

---

7. As an illustration of a relative-mean-income curve, consider the Pareto distribution with $F(y) = 1 - y^{-\alpha}$ and $\alpha > 1$. Then $\mu = \alpha/(\alpha-1)$, and $\xi_{p_i} = (1-p_i)^{-1/\alpha}$, so that the relative-mean-income-curve ordinates are $\xi_{p_i}/\mu = (\alpha-1/\alpha)(1-p_i)^{-1/\alpha}$. Thus for selected upper-tail values of p_i and alternative values of α, the corresponding relative-mean-income ordinates are easily computed.

p_i =		.7	.8	.9	.95
α = 1.5	-	.7438	.9746	1.5474	2.4562
2.0	-	.9129	1.1181	1.5813	2.2364
2.5	-	.9712	1.1422	1.5072	1.9887
3.0	-	.9958	1.1400	1.4362	1.8097
4.0	-	1.0134	1.1216	1.3338	1.5860

and their corresponding Lorenz curve derivatives.[8] Consequently, an alternative way of saying that it is useful for an applied distribution analyst to provide a set of income quantiles to go with an estimated Lorenz curve is that he should provide an estimated relative-mean-income curve as well, as done, for example, in some work of Beach et. al. [6]. A relative-mean-income curve thus has an important inferential role in applied work as well as a useful descriptive role in distribution analysis.

Note also the relatively simple structure of the asymptotic covariance matrix in (3.3). For positive incomes, V_L has all positive elements; that is, between cumulative income shares, covariances are quite reasonably positive. As one moves down the principal diagonal of terms $(\xi_{p_i}/\mu)^2$ $p_i(1-p_i)$, the component $p_i(1-p_i)$ increases to a maximum at the median value $p_i = .5$ and then decreases, while the square of the relative-mean-income value increases steadily from $(\xi_{p_1}/\mu)^2$ to $(\xi_{p_K}/\mu)^2$. Thus the variances increase over the range p_i to beyond the median and then may either increase

---

8. This should not be at all surprising since we know that (i) the proportions $F(\hat{\xi}_{p_i})$ and $F(\hat{\xi}_{p_j})$ for $i < j$ are asymptotically normal with asymptotic covariance $p_i(1-p_j)/N$ [33, p. 271], and that (ii) the derivative of the function $\Phi(F(\xi_{p_i}))$ is $d\Phi(\xi_{p_i})/dF = \xi_{p_i}/\mu$. Consequently, the income share functions $\Phi(F(\hat{\xi}_{p_i}))$ and $\Phi(F(\hat{\xi}_{p_j}))$ are also asymptotically normal with asymptotic covariance $(\xi_{p_i}/\mu)\ (\xi_{p_j}/\mu)\ p_i\ (1-p_j)/N$.

or decrease depending on which effect dominates.[9] Typically, for skewed distributions of income or wealth, the estimated variances have been found to reach a maximum in the interval between p = .70 and p = .85 and thereafter decline. Also note that the asymptotic squared correlation coefficient between cumulative shares corresponding to p_i and p_j $(p_i < p_j)$ is $p_i(1-p_j)/p_j(1-p_i)$. That is, the correlations are independent even of the quantile levels and depend solely on the (known) abscissa proportions p_i, p_j. As one moves along the minor diagonal of V_L where $p_i + p_j = 1$, the correlation is maximized at the median when i = j and minimized at the two ends of the diagonal where asy. $\text{cor}^2(\hat{\Phi}_i, \hat{\Phi}_j) = p_1^2/p_K^2$.

III.2 Asymptotic Distribution of Income Shares

The line of argument to derive the asymptotic distribution of Lorenz curve ordinates holds also for a set of income shares. If the Lorenz curve ordinates represent cumulative income shares, the differences between successive ordinates corresponding to adjacent quantiles represent income shares between different quantiles. If there are K quantiles (e.g., K = 9 in the case of deciles), then there are K + 1 (population) quantile shares

$$\psi_i = \Phi(\xi_{p_i}) - \Phi(\xi_{p_{i-1}}) \qquad i = 1, 2, \ldots, K+1 \qquad (3.5)$$

where we set $\Phi(\xi_{p_0}) = 0$ and $\Phi(\xi_{p_{K+1}}) = 1$. Since $\hat{\psi}_i = \hat{\Phi}_i - \hat{\Phi}_{i-1}$ is just a

---

9. In the case of the Pareto distribution with $F(y) = 1 - y^{-\alpha}$ for $\alpha > 1$, the asymptotic variance is

$$\frac{v_{ii}}{N} = \left(\frac{1}{N}\right)\left(\frac{\alpha-1}{\alpha}\right)^2 p_i(1-p_i)^{\frac{\alpha-2}{\alpha}}.$$

For given N and α, this is maximized at

$$p^* = \frac{1}{2}\left(\frac{\alpha}{\alpha-1}\right).$$

Consequently, when α = 2, 2.5, and 3, p^* = 1.0, .8333, and .75 respectively.

difference in sample Lorenz curve ordinates which are asymptotically normal with asymptotic covariance matrix $(1/N)V_L$, it is clear that the sample income share statistics are also asymptotically (K+1) - variate normal with asymptotic mean $\psi = (\psi_1, \psi_2, \ldots, \psi_{K+1})'$ and asymptotic covariance matrix $(1/N)V_s$ where $V_s = J_s V_L J_s'$ and the (K+1) x K gradient matrix

$$J_s = \left[\frac{\partial\psi_i}{\partial\Phi_j}\right] = \begin{bmatrix} 1 & & & & 0 \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & 0 & & -1 & 1 \\ & & & & -1 \end{bmatrix} \tag{3.6}$$

Thus combining (3.3) and (3.6), one can check that the ij'th element of the symmetric matrix V_s where $1 \leq i \leq j \leq K+1$ is equal to

$$v_{ij}^s = (1/\mu^2)[\xi_{p_{i-1}} \xi_{p_{j-1}} (1-p_{j-1}) - \xi_{p_i} \xi_{p_{j-1}} p_i (1-p_{j-1})$$

$$- \xi_{p_{i-1}} \xi_{p_j} p_{i-1} (1-p_j) + \xi_{p_i} \xi_{p_j} p_i(1-p_j)] \tag{3.7}$$

where $p_0 = 0$, $p_{K+1} = 1$, $\xi_{p_0} = 0$, and $\xi_{p_{K+1}}$ is assumed finite.

Again, it is evident that V_s does not depend upon the underlying population density function $f(\cdot)$, so that model-free inferences concerning income shares are again feasible. Note also that, in contrast to V_L, V_s is of dimension (K+1) x (K+1) and singular since the sum of the K+1 income shares is identically one.

In order to compute (asymptotic) standard errors for income shares, one simplifies (3.7) by setting $i = j$ to

$$v_{ii}^s = (1/\mu^2) [\xi_{p_{i-1}}^2 p_{i-1}(1-p_{i-1}) - 2\xi_{p_i}\xi_{p_{i-1}} p_{i-1}(1-p_i)$$
$$+ \xi_{p_i}^2 p_i (1-p_i)] \tag{3.8}$$

It is then immediately evident from (3.8) that, to compute standard errors for income shares, one need compute only 2K-1 elements -- the K diagonal elements and K-1 first-superdiagonal elements -- of the V_L matrix and not the full set of K(K-1)/2 different elements in V_L. The (asymptotic) standard error for the i'th income share $\hat{\psi}_i$ can thus be computed as

$$[(1/N\bar{Y}^2)\ [\hat{\xi}^2_{p_{i-1}}\ p_{i-1}(1-p_{i-1}) - 2\ \hat{\xi}_{p_i}\ \hat{\xi}_{p_{i-1}}\ p_{i-1}(1-p_i) + \hat{\xi}^2_{p_i}\ p_i(1-p_i)]^{\frac{1}{2}} \tag{3.9}$$

The asymptotic variances of bottom and top income shares are particularly easy to compute. The share statistic for the lowest 100 p_i% of the sample is simply $\hat{\psi}_i = \hat{\Phi}_i$ which has the (asymptotic) standard error $(\frac{\hat{\xi}_{p_i}}{\bar{Y}})\sqrt{\frac{p_i(1-p_i)}{N}}$. The share statistic for the top 100(1-p_i)% is $\hat{\psi}_i = 1 - \hat{\Phi}_i$, so that the corresponding (asymptotic) standard error is also $(\frac{\hat{\xi}_{p_i}}{\bar{Y}})\sqrt{\frac{p_i(1-p_i)}{N}}$.

IV. Standard Errors for Gini Coefficients

A corollary of deriving the asymptotic distribution of sample Lorenz curve ordinates is that one can also do so for an interpolated approximation to the Gini coefficient, perhaps the single most frequently used summary measure of income inequality in a distribution. While Gastwirth [16] and Kakwani [18] have derived asymptotic distributions for estimates of various other summary inequality measures, this appears to be the first such derivation for the Gini coefficient. The approach again is model-free, and does not require *a priori* specification of the underlying parent distribution such as is involved in maximum likelihood methods used by Kakwani [18]. The geometric approach used here also avoids the rather substantial difficulties of the perhaps more natural approach [21, p. 241] of first examining the distribution of the mean absolute difference,

$$\Delta = \int_0^\infty \int_0^\infty | y_1 - y_2 | dF(y_1)dF(y_2),$$

which appears in the numerator of the Gini coefficient.

The (population) Gini coefficient of concentration, Γ, lying in the interval (0,1) for positive incomes, is geometrically equal to twice the area between the Lorenz curve and the absolute equality diagonal [21, p. 49]. If one interpolates linearly along the Lorenz curve between adjacent quantile ordinates and uses a trapezoidal integration formula, the Gini coefficient[10] may be estimated as

$$\hat{\Gamma} = G = (1/K+1) \Sigma_{i=1}^{K+1} (p_i - \hat{\Phi}_i + p_{i-1} - \hat{\Phi}_{i-1}) \tag{4.1}$$

if the p_i's are equally spaced. Therefore the (Kx1) gradient vector for the linear transformation (4.1) is $j = (-2/(K+1), \ldots, -2/(K+1))'$, and one obtains from the results of Section III.1 that $\sqrt{N}(G-\Gamma)$ also has a limiting normal distribution with mean zero and variance

$$j'V_L\, j = [4/(K+1)^2]\, \Sigma_{i=1}^{K} \Sigma_{j=1}^{K} v_{ij}^L,$$

10. Note that this is the only point at which interpolation has been used in this paper. The expression for the estimated variance of G is thus approximate in that it reflects both sampling errors as well as interpolation errors. One could if one wished also use an alternative interpolation formula such as Gastwirth's [15] "upper-bound" interpolation rule or some rule-of-thumb combination of the two.

where the summation is over all elements of the V_L covariance matrix. The corresponding (asymptotic) standard error of G is thus

$$\text{S.E.}(G) = \frac{2}{(K+1)} \left[\frac{\Sigma_i \Sigma_j \hat{v}^L_{ij}}{N}\right]^{\frac{1}{2}}, \tag{4.2}$$

where $\hat{v}^L_{ij} = (\hat{\xi}_{p_i}/\bar{Y})\ (\hat{\xi}_{p_j}/\bar{Y})\ p_i\ (1-p_j)$ for $i < j$.

Since the Gini coefficient is expressed as a function of the Lorenz curve ordinates for given p_i's, it too has the property of allowing model-free statistical inference. The relative-mean-deviation inequality statistic, in contrast, does not (Beach [5]). However, the estimated coefficient and its standard error do depend on the coarseness of the interpolation intervals $[p_i, p_{i-1}]$, so that it is advisable when reporting inference results based on (4.1) and (4.2) to indicate also the interval size (e.g., deciles or quintiles) used in the interpolation.

V. Hypothesis Testing with Quantile Results

V.1) Hypothesis Tests on Income Shares

Given the asymptotic distribution results on estimated income shares derived in the last section, one is now able to consider directly the problem of hypothesis testing with income shares.

i) Tests on Single Share Statistics

First of all, consider the case where there is some hypothesized value ψ_i^0 to which the sample share statistic, $\hat{\psi}_i$ is being compared (for example, that the bottom 10% of recipients get only 5% of total income).

From the results of Section III.2, it is clear that the appropriate test statistic under H_0: $\psi_i = \psi_i^0$ is $z_i = (\hat{\psi}_i - \psi_i^0)/(\hat{v}_{ii}^s/N)^{\frac{1}{2}}$, which is to be compared to the critical values on a standard normal table for a specified level of significance α.

More typically, however, the distribution analyst is more interested in comparing income shares between two alternative distributions (for example, between two time periods or two regions). Specifically, suppose one has two corresponding income share statistics $\hat{\psi}_{1i}$ and $\hat{\psi}_{2i}$ based respectively on two independent samples of sizes N_1 and N_2. According to a null hypothesis, H_0: $\psi_{1i} = \psi_{2i}$ against, say, H_1: $\psi_{1i} \neq \psi_{2i}$ for a given particular quantile share. Under the independence assumption, the appropriate standard normal test statistic becomes $z_i = (\hat{\psi}_{1i} - \hat{\psi}_{2i})/[(\hat{v}_{ii}^{s_1}/N_1) + (\hat{v}_{ii}^{s_2}/N_2)]^{\frac{1}{2}}$ where $\hat{v}_{ii}^{s_1}$ and $\hat{v}_{ii}^{s_2}$ are the estimated variances based on (3.8) for samples 1 and 2 respectively.

Tests on single share statistics such as just considered are most likely to be appropriate when looking at either top or bottom shares in a distribution.[11]

ii) Joint Test on a Set of Income Shares

When evaluating an overall distribution of income, one may be more concerned with a set of income shares. For purposes of exposition, suppose

---

11. It may be remarked that standard "t-ratios" typically reported for individual regression coefficients are not so interesting for estimated share statistics. Perhaps the more appropriate "standard" on which to base individual test statistics is the null hypothesis of absolute equality. Consequently, instead of reporting individual "t-ratios", $t = \hat{\psi}_i/\sqrt{\hat{v}_{ii}^s/N}$, it may be more appropriate to report individual "z-ratios", $z = (\hat{\psi}_i - p_i)/\sqrt{\hat{v}_{ii}^s/N}$.

one is interested in the full set of K quantile share statistics (one share statistic, say the last, is omitted as being linearly dependent on the others). For example, one may have a model of income-generating behavior as in Fair (1971) and wish to compare an actual distribution of income shares, say $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_K)'$, to an hypothesized set of income shares $\psi^0 = (\psi_1^0, \psi_2^0, \ldots, \psi_K^0)'$ specified by the theoretical model. In this case, one wishes to test H_o: $\psi = \psi^0$ against the uninformative alternative H_1: $\psi \neq \psi^0$. From the results of Section III.2, under the null hypothesis, $\sqrt{N}(\hat{\psi} - \psi^0)$ is asymptotically distributed as a K-variate normal with mean zero and covariance matrix $\bar{V}_s$, where the bar notation on V_s indicates that the last row and column of the V_s matrix have been deleted. Consequently, the test statistic

$$c_1 = N(\hat{\psi} - \psi^0)' \, \hat{\bar{V}}_s^{-1} \, (\hat{\psi} - \psi^0) \tag{5.1}$$

is asymptotically distributed under H_o as a (central) chi-squared variate with K degrees of freedom.

It should be remarked, however, that the actual computations involved in the income share test (5.1) (and in subsequent tests as well) are much simpler than may first appear as there is no need to invert the matrix $\hat{\bar{V}}_s$ numerically. If the K x K) nonsingular matrix $\bar{J}_s$ is defined as

$$\bar{J}_s = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix},$$

it can be seen that the share covariance matrix

$$\bar{V}_s = \bar{J}_s V_L \bar{J}_s' ,$$

so that

$$(\bar{V}_s)^{-1} = (\bar{J}_s')^{-1} V_L^{-1} (\bar{J}_s)^{-1},$$

where it can also be checked that

$$(\bar{J}_s)^{-1} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \cdots & 1 \end{bmatrix} .$$

Thus any arbitrary quadratic form in the matrix $(\bar{V}_s)^{-1}$ can be written as

$$a'(\bar{V}_s)^{-1}a = b'V_L^{-1}b, \tag{5.2}$$

where

$$b=(\bar{J}_s)^{-1}a = \begin{bmatrix} a_1 \\ a_1+a_2 \\ a_1+a_2+a_3 \\ \vdots \\ a_1+a_2+ \cdots +a_K \end{bmatrix}, \tag{5.3}$$

so that it becomes now a quadratic form in the matrix V_L^{-1}, the inverse of the Lorenz curve (asymptotic) covariance matrix.

V_L, however, can be shown to have a simple analytic inverse. Specifically, it will be recalled that V_L = RAR where R is a diagonal matrix and A = P - pp' from (3.3). Now the matrix A^{-1} can be seen to have a very simple structure, with elements

$$a^{ii} = \frac{p_{i+1} - p_{i-1}}{(p_{i+1}-p_i)(p_i-p_{i-1})} \qquad \text{for } i = 1, \ldots, K, \qquad (5.4a)$$

$$a^{i,i+1} = a^{i+1,i} = \frac{-1}{(p_{i+1}-p_i)} \qquad \text{for } i = 1, \ldots, K-1, \qquad (5.4b)$$

and zeros elsewhere [26, p. 385]. Again, for convenience, set $p_0 = 0$ and $p_{K+1} = 1$. Consequently, any quadratic form in the matrix V_L^{-1} can be written as

$$b'V_L^{-1}b = \Sigma_{i=1}^{K} \frac{(p_{i+1}-p_{i-1})}{(p_{i+1}-p_i)(p_i-p_{i-1})} \; b_i^2 - 2\,\Sigma_{i=2}^{K} \frac{b_i b_{i-1}}{(p_i-p_{i-1})} \,. \qquad (5.5)$$

Thus one needs to compute only 2K-1 terms in (5.5) instead of inverting a (KxK) matrix numerically. When one is working with deciles or vigintiles, for examples, this is a substantial computational reduction. The test statistic in (5.1) can thus be re-expressed as

$$c_1 = N[\Sigma_{i=1}^{K} \frac{(p_{i+1}-p_{i-1})}{(p_{i+1}-p_i)(p_i-p_{i-1})} \left(\frac{\hat{\xi}_{p_i}}{\hat{\mu}}\right)^{-2} b_i^2$$

$$- 2\,\Sigma_{i=2}^{K} \frac{b_i\; b_{i-1}}{(p_i-p_{i-1})} \left(\frac{\hat{\xi}_{p_i}}{\hat{\mu}}\right)^{-1} \left(\frac{\hat{\xi}_{p_{i-1}}}{\hat{\mu}}\right)^{-1}], \qquad (5.6a)$$

where $b_i = \Sigma_{j=1}^{i} (\hat{\psi}_j - \psi_j^0)$.

Clearly, one could also work out an intermediate case where a test is performed on a set of only L quantile shares where $1 \leq L \leq K$ based on an asymptotic chi-squared statistic with L degrees of freedom.

iii) Joint Test of a Difference of Two Independent Sets of Income Shares

When one is comparing alternative distributions, however, one may be more concerned with testing for differences in sets of share statistics between two sample distributions corresponding, for example, to different periods or different regions. Specifically, suppose one distribution is characterized by a set of K quantile shares $\hat{\psi}_1 = (\hat{\psi}_{11}, \hat{\psi}_{12}, \ldots, \hat{\psi}_{1K})'$ and the second by $\hat{\psi}_2 = (\hat{\psi}_{21}, \hat{\psi}_{22}, \ldots, \hat{\psi}_{2K})'$ and the samples are drawn independently of size N_1 and N_2 respectively. The null hypothesis one may wish to test then is

$$H_0: \psi_1 = \psi_2 \text{ against } H_1: \psi_1 \neq \psi_2.$$

Now the two share-covariance matrices V_{s1} and V_{s2} can be seen to be equal if and only if $(\xi_{1p_i}/\mu_1) = (\xi_{2p_i}/\mu_2)$ for all i; that is, if the relative mean income curves are the same for the two distributions. But if the relative mean income curves are the same, so also are the corresponding Lorenz curves, and the corresponding sets of quantile share statistics. Consequently, under the null hypothesis that $\psi_1 = \psi_2$, we shall also assume that the two covariance matrices are equal, $V_{s1} = V_{s2} = V_s$.

Under the null hypothesis, then, one can see that the vector difference $(\hat{\psi}_1 - \hat{\psi}_2)$ is asymptotically K-variate normal with mean zero and covariance matrix $(1/N_1 + 1/N_2)\ \bar{V}_s$. Consequently, an appropriate test statistic for H_0 is

$$c_2 = \left(\frac{N_1 N_2}{N_1+N_2}\right) (\hat{\psi}_1 - \hat{\psi}_2)'\ \hat{\bar{V}}^{-1}\ (\hat{\psi}_1 - \hat{\psi}_2) \qquad (5.7)$$

which will also be asymptotically chi-squared with K degrees of freedom.[12]

---

12. Since covariance matrices are assumed to be the same in the two samples, estimates of the elements of $\bar{V}_s$ should be based on a combined sample. A convenient approximation to the combined relative-mean-income ordinates, however, may be provided simply by the weighted average

$$\frac{\hat{\xi}_i}{\hat{\mu}} = \left(\frac{N_1}{N_1+N_2}\right) \left(\frac{\hat{\xi}_{1p_i}}{\hat{\mu}_1}\right) + \left(\frac{N_2}{N_1+N_2}\right) \left(\frac{\hat{\xi}_{2p_i}}{\hat{\mu}_2}\right).$$

Following the same argument presented for c_1, one can alternatively and more simply compute c_2 by the formula (5.6a) where now

$$b_i = \Sigma_{j=1}^{i} (\hat{\psi}_{1j} - \hat{\psi}_{2j}). \tag{5.7b}$$

Again one can also formulate joint tests for differences in subsets of quantile shares as well.

V.2) Hypothesis Tests and Confidence Bands on Lorenz Curves

In the case of Lorenz curves, tests of individual ordinates are not typically of much concern, so that we consider only joint tests on the full set of K Lorenz curve ordinates analogous to those just discussed for income shares.

i) Joint Tests on Lorenz Curve Ordinates

Since much of the framework for hypothesis testing of Lorenz curve ordinates has already been laid out, the present discussion can be fairly brief. To compare a hypothetical or theoretical Lorenz curve $\Phi^0 = (\Phi_1^0, \Phi_2^0, \ldots, \Phi_K^0)'$ against an empirically estimated curve $\hat{\Phi} = (\hat{\Phi}_1, \hat{\Phi}_2, \ldots, \hat{\Phi}_K)'$ in order to test H_0: $\Phi = \Phi^0$ vs H_1: $\Phi \neq \Phi^0$, one can again use an asymptotic chi-squared test statistic

$$c_3 = N(\hat{\Phi} - \Phi^0)' \hat{V}_L^{-1} (\hat{\Phi} - \Phi^0) \tag{5.8}$$

with K degrees of freedom. To compare two separate Lorenz curve estimates $\hat{\Phi}_1$ and $\hat{\Phi}_2$ from independent samples, in order to test H_0: $\Phi_1 = \Phi_2$ vs H_1: $\Phi_1 \neq \Phi_2$, one can use the statistic

$$c_4 = \left(\frac{N_1 N_2}{N_1+N_2}\right) (\hat{\phi}_1 - \hat{\phi}_2)' \hat{V}_L^{-1} (\hat{\phi}_1 - \hat{\phi}_2) \tag{5.9}$$

which is also asymptotically chi-squared with K degrees of freedom under the null hypothesis and accompanying assumption of equal variances.

Just as the share test statistics can be computed without having to invert numerically a (KxK) covariance matrix, so also can c_3 and c_4. Specifically, using the result in (5.5), one can re-express (5.8) as

$$c_3 = N[\Sigma_{i=1}^K \frac{(p_{i+1}-p_{i-1})}{(p_{i+1}-p_i)(p_i-p_{i-1})} \left(\frac{\hat{\xi}_{p_i}}{\hat{\mu}}\right)^{-2} (\hat{\Phi}_i - \Phi_i^0)^2$$

$$-2\Sigma_{i=2}^K \frac{1}{(p_i-p_{i-1})} \left(\frac{\hat{\xi}_{p_i}}{\hat{\mu}}\right)^{-1} \left(\frac{\hat{\xi}_{p_{i-1}}}{\hat{\mu}}\right)^{-1} (\hat{\Phi}_i - \Phi_i^0)(\hat{\Phi}_{i-1}-\Phi_{i-1}^0)] \tag{5.10}$$

and (5.9) as

$$c_4 = \left(\frac{N_1 N_2}{N_1+N_2}\right)[\Sigma_{i=1}^K \frac{(p_{i+1}-p_{i-1})}{(p_{i+1}-p_i)(p_i-p_{i-1})} \left(\frac{\hat{\xi}_{p_i}}{\hat{\mu}}\right)^{-2} (\hat{\Phi}_{1i}-\hat{\Phi}_{2i})^2 \tag{5.11}$$

$$- 2\Sigma_{i=2}^K \frac{1}{(p_i-p_{i-1})} \left(\frac{\hat{\xi}_{p_i}}{\hat{\mu}}\right)^{-1} \left(\frac{\hat{\xi}_{p_{i-1}}}{\hat{\mu}}\right)^{-1} (\hat{\Phi}_{1i}-\hat{\Phi}_{2i})(\hat{\Phi}_{1i-1}-\hat{\Phi}_{2i-1})].$$

One particularly interesting problem where one may wish to apply the above inference procedures is that of statistically testing Atkinson's [2] distributional ranking criterion involving Lorenz curves. Specifically, one may wish to use the criterion of nonintersecting Lorenz curves to define a ranking or comparison of inequality between two distributions (as opposed to defining a ranking of distributions per se), as applied, for example, in Beach et.al. [6]. To test empirically the hypothesis of one Lorenz curve lying statistically significantly inside another, one may start from a situation of one estimated Lorenz curve $\hat{\Phi}_1$ indeed lying uniformly above another $\hat{\Phi}_2$ (i.e.: $\hat{\Phi}_{1i} > \hat{\Phi}_{2i}$ for all i = 1, ..., K), and then use statistic c_4 to test H_0: $\Phi_1 = \Phi_2$ against the one-sided alternative H_1: $\Phi_1 > \Phi_2$.

ii) <u>Confidence Band for Lorenz Curves</u>

Along with the hypothesis tests so far described, it would be desirable from a graphical point of view to supplement an illustrated Lorenz curve with some kind of confidence band about it over its full length. One could then immediately see graphically how accurately the illustrated Lorenz curve has been estimated, and particularly over some regions more tightly than others.

Perhaps an initial approach to this problem might simply be to construct a band of, say, two standard errors of $\hat{\Phi}_i$ on both sides of the estimated Lorenz curve ordinates. While such a band may have some descriptive interest in illustrating the <u>relative</u> widths of individual ordinate confidence intervals, it is not a very useful analytical device because it treats individual ordinate estimates as separate and unrelated. What is wanted instead is a joint confidence band or set of simultaneous confidence intervals that incorporate the market interdependence of the individual ordinate estimates for Lorenz curves. As is well known in the statistical literature this is the classical problem of determining a set of simultaneous confidence intervals or multiple comparisons for a given joint level of confidence, and there is no unique way of handling the problem. Perhaps the best known approach is Scheffé's [29, pp. 68-70] projection method. If $d_\alpha = \sqrt{\chi_K^2}$ is the square root of the 100 (1-α)% critical value on a chi-square distribution with K degrees of freedom, then the probability is at least 100(1-α)% that the K intervals $(\hat{\phi}_i - d_\alpha\sqrt{v_{ii}^L}, \hat{\Phi}_i + d_\alpha\sqrt{v_{ii}^L})$ jointly contain the K population ordinates $\Phi_i, \Phi_2, \ldots, \Phi_K$. Consequently, an approximate set of simultaneous confidence intervals is provided by a band of d_α standard errors in width on both sides of the estimated Lorenz curve ordinates. In the case of decile ordinates (K=9) with α = .05, the corresponding value of d_α is $d_\alpha = \sqrt{16.919} = 4.11$. This compares with the two-standard-errors rule that corresponds to treating the ordinates as separate and unrelated.

Alternative approaches to the simultaneous confidence interval problem are also available [30, pp. 126-132]. Bonferroni t-intervals, for example, are based on the critical value of $t_\upsilon^{\alpha/2K}$ for d_α from the t-distribution with υ degrees of freedom. Asymptotically, one may simply use $z^{\alpha/2K}$ from the standard normal distribution for large micro data samples. In the above case where α = .05 and K = 9, the Bonferroni critical value is d_α = 2.78 which is substantially smaller than that obtained from the Scheffé procedure, and consequently in this case to be preferred.

VI. Illustrative Empirical Results

Several of the tests of Section V are now illustrated with two sources of micro data, one for the United States from Danziger and Taussig [9],[13] and one for Canada from Beach et al. [6].

Table 1 provides the background data on decile income levels, decile shares, and Lorenz curve ordinates for United States census unit households (reporting positive income) from the CPS for the two years 1967 and 1976.

---

13. The author would like to thank Prof. Sheldon Danziger for providing the data in Table 1.

These estimates are based on very large data sets (N_1 = 48,191 for 1967 and N_2 = 58,063), and appear roughly similar except for the inflation of income values over the period; as a result, the sample mean increased from \$7692 in 1967 to \$14,087 in 1976.

Table II provides (asymptotic) standard errors on the decile income shares as computed by (3.9) (given in percents) for the two years, and also z-statistics on the difference of individual shares, $\hat{\psi}_{1i} - \hat{\psi}_{2i}$. Judging the shares separately, one can see the differences are individually significantly different from zero in the first, third, fourth, fifth, and eighth decile shares on conventional significance levels. Note also how the standard errors are consistently slightly smaller for 1976 because of the larger sample size.

Table III provides more summary test statistics for differences in overall inequality between the two years. A joint test of the difference between the two Lorenz curves is computed from (5.11) to be C_4 = 54.46 which is seen to be highly significant at any conventional levels of significance. The Gini coefficient standard errors are also computed (based on deciles) and yield test statistics for significant difference from zero (i.e., absolute equality) of 163. and 183. for 1967 and 1976 respectively. However, the difference between the two Ginis has a z-ratio of only -2.091 which lies between a 95% and 99% confidence-level cut-off on the normal table with a two-tailed test. Thus it is quite clear that a test on Gini coefficients is not at all equivalent to a test on significant differences in the overall Lorenz curve. In the first place, one is a single test, while the other a joint test. Secondly, one has an assumed aggregation structure and implicit

TABLE I

Decile Incomes, Shares, and Lorenz Curves: 1967, 1976

1) United States CPS Households, 1967

Decile ($10p_i$)	Decile Level ($\hat{\xi}_{p_i}$)	Decile Share ($100\hat{\psi}_i$)	Lorenz C. Ord. ($100\hat{\Phi}_i$)
1	$1,441	1.00%	1.00%
2	2,700	2.66	3.66
3	4,056	4.38	8.04
4	5,457	6.24	14.28
5	6,750	7.95	22.23
6	8,000	9.59	31.82
7	9,504	11.34	43.16
8	11,390	13.55	56.71
9	14,500	16.57	73.28
10		26.72	100.00
G_1 = .3992		$\bar{Y}_1$ = $7,692	N_1 = 48,191

2) United States CPS Households, 1976

Decile ($10p_i$)	Decile Level ($\hat{\xi}_{p_i}$)	Decile Share ($100\hat{\psi}_i$)	Lorenz C. Ord. ($100\hat{\Phi}_i$)
1	$2,935	1.16%	1.16%
2	4,875	2.73	3.89
3	7,000	4.18	8.07
4	9,285	5.78	13.85
5	11,870	7.50	21.35
6	14,580	9.36	30.71
7	17,540	11.37	42.08
8	21,350	14.10	56.18
9	27,450	17.01	73.19
10		26.80	100.0
G_2 = .4061		$\bar{Y}_2$ = $14,087	N_2 = 58,063

Source: See footnote 13, and Danziger and Taussig [9].

TABLE II

Decile Shares and Standard Errors: United States 1967 and 1976

Decile	1967	1976	z dif.
1	1.00% (0.026)	1.16% (0.026)	-4.40*
2	2.66 (0.051)	2.73 (0.045)	-1.04
3	4.38 (0.074)	4.18 (0.036)	2.06*
4	6.24 (0.096)	5.78 (0.081)	3.66*
5	7.95 (0.116)	7.50 (0.101)	2.93*
6	9.59 (0.134)	9.36 (0.121)	1.27
7	11.34 (0.156)	11.37 (0.143)	-0.14
8	13.55 (0.182)	14.10 (0.168)	-2.20*
9	16.57 (0.215)	17.01 (0.202)	-1.49
10	26.72 (0.258)	26.80 (0.243)	-0.23

*denotes significantly different from zero on the basis of a two-tailed test of a standard normal variate with $\alpha = .05$.

Source: Based on data in Table I.

TABLE III

Summary Test Statistics: United States 1967 and 1976

	1967	1976
Gini Coefficient:	.3992 (.00245)z=162.9	.4061 (.00222)z=182.9
Lorenz Curve Difference:	$c_4 = 54.46 > \chi^2_9 = 23.59$ at $\alpha = .005$	
Gini Coef. Difference:	$d = G_{67} - G_{76} = -.0069$	
	S.E.(d) = .00330	
	zd = -2.091	
	[z(α=.05) = 1.960, z(α=.01) = 2.326].	

Source: Based on data in Table I.

social welfare function built into it while the other does not. In the case of two intersecting Lorenz curves, for example, the corresponding Gini coefficients can be the same while the Lorenz curves are quite different. In general, the Lorenz curve joint test is to be preferred to that on the Gini coefficient as a less restrictive test.

It can also be seen that with such large sample sizes, even rather similar looking distributions can be quite sharply distinguished as to their relative structure of inequality. At the same time, the size of "sampling error" is on the low side relative to "interpolation error" as found, for example, by Gastwirth [15], who computed interpolation error bounds on the Gini coefficients for the 1967 CPS data with 10 income groups. The width of the interval between upper and lower interpolation bounds for three different interpolation procedures was calculated as .020, .019, and .009. These may be compared to an approximate 95% confidence interval on G for 1967 of ±2 standard errors or an interval width of .009.

Finally, Table IV provides Lorenz curve data on family total income for all (census) family units in the province of Ontario, Canada, for 1973 taken from a recent empirical study by the author and others (Beach et al. [6]) and computed based on a vigintile (K+1=20) income disaggregation and a sample size of 7624 family units. This finer level of disaggregation shows the Lorenz curve standard errors increasing up until the sixteenth vigintile and then decreasing in size. The third column provides joint confidence intervals for the nineteen vigintile ordinates based on "Bonferroni-z" intervals. At a 95% level of confidence, the asymptotic Bonferroni-z value for d_α is 3.01 (Seber (1977), p. 131) compared to the corresponding

TABLE IV

Lorenz Curve Vigintile Ordinates
Family Total Income for All Family Units
Ontario, 1973

Vigintile	Pt. Est.	Est.±3.01 S.E.	S.E.
1	0.39%	0.29 - 0.49%	0.034%
2	1.23	1.03 - 1.43	0.067
3	2.37	2.05 - 2.70	0.108
4	3.95	3.45 - 4.45	0.166
5	5.98	5.31 - 6.65	0.223
6	8.47	7.61 - 9.33	0.287
7	11.47	10.39 - 12.55	0.358
8	14.97	13.71 - 16.23	0.417
9	18.92	17.49 - 20.35	0.475
10	23.32	21.73 - 24.91	0.529
11	28.14	26.42 - 29.86	0.572
12	33.38	31.53 - 35.23	0.614
13	39.07	37.13 - 41.01	0.644
14	45.21	43.20 - 47.22	0.668
15	51.83	49.78 - 53.88	0.682
16	59.00	56.94 - 61.06	0.684
17	66.83	64.81 - 68.85	0.670
18	75.51	73.59 - 77.43	0.638
19(K)	85.63	83.95 - 87.31	0.558

$\bar{Y}$ = \$11,091 G = .374

N = 7624. (.00639) z = 58.5

Source: Beach et al. [6], Tables 9.1, 9.4, and 9.5.

asymptotic Scheffé value for d_α which would be $d_\alpha = \sqrt{\chi^2_{19}} = 5.49$. Consequently, the narrower Bonferroni intervals have been used in the table.

Finally, one may remark on the substantially larger standard error for the estimated Gini coefficient in Table IV than in Table III because of the smaller sample size on which it is based. It has also been computed from vigintile values, whereas the earlier figures were based on decile values. However, if one recomputed the standard error in Table IV in more aggregated fashion, one would obtain values of .006335 from decile figures and .006160 from quintile figures compared to .006391 from the reported vigintile figures. That is, the Gini standard errors appear quite insensitive to the level of aggregation used and differ less than 4% between quintile and vigintile levels of disaggregation.

VII. Review and Conclusions

The general objective of this paper has been to extend the standard techniques of statistical inference to applied income distribution work at a disaggregated level of analysis. Sections II-IV of the paper introduced the essential background material on the asymptotic distributions of income quantiles, and then used them to derive model-free standard errors and confidence intervals for income share statistics, Lorenz curve estimates, and estimated Gini coefficients. The only additional information required to estimate the asymptotic covariance matrices involved is that of a relative mean income curve. Sections V and VI then provided several hypothesis tests on income shares and Lorenz curves which are typically of most interest to applied distribution analysts.

Three general conclusions emerge from this paper. First, it clearly follows that model-free statistical inference on Lorenz curves, income shares, and Gini coefficients is both feasible and remarkedly simply to carry out. Consequently, it is hoped that henceforth applied distribution analysis will be carried on in the framework of standard statistical inference. Second, when an analyst is reporting his empirical results in terms of Lorenz curves, he should also report estimated relative-mean-income ordinates so as to allow a reader to carry out inferences on the Lorenz curve figures. Third, statistical agencies providing published distribution data should also include, along with income share and histogram data, quantile income level estimates such as decile levels which researchers can then use for statistical inference purposes.

Appendix

Lemma 2: Under the conditions of Theorem 1, if the population density has finite mean and variance, $\sqrt{N}(\hat{\Phi}_i - \Phi_i)$ and $\sqrt{N}(\Phi(\hat{\xi}_{p_i})-\Phi_i)$ have the same limiting distribution.

Proof: The first part of the proof is a modification of the arguments in Gastwirth's [16] Theorem 1.

Recall, first of all, that by the Central Limit Theorem $z = N^{\frac{1}{2}}(\bar{Y} - \mu)/\sigma$ has an asymptotic standard normal distribution if the Y's are drawn (as assumed) from a random sample. Also by Theorem 1 of the text,

$$\varepsilon = N^{\frac{1}{2}}(\hat{\xi}_{p_i} - \xi_{p_i})f(\xi_{p_i})/[p_i(1-p_i)]^{\frac{1}{2}} \tag{A1}$$

is asymptotically standard normal as well.

Now in order to transform a conditional mean problem into an unconditional mean problem, introduce the random variable

$$\begin{aligned} I_j^i &= 1 \text{ if } Y_j < \xi_{p_i} \\ &= 0 \text{ otherwise} \end{aligned} \tag{A2}$$

where Y_j denotes the j'th observation in the random sample drawn from the continuous density $f(\cdot)$ with finite mean and variance. The number of observations less than ξ_{p_i} is a binomial random variable with parameters N and p_i, and

$$\tau_i = E(I_j^i Y_j) = \int_0^{\xi_{p_i}} y \; dF(y) = p_i \; E(Y_j | Y_j \leq \xi_{p_i}).$$

Consider then the asymptotic distribution of the conditional mean estimator $\bar{Y}_{\hat{\xi}_{p_i}} = (1/n_i) \Sigma_{Y_j \leq \hat{\xi}_{p_i}} Y_j$ where $n_i = [Np_i]$. Let

$$S_i = \sqrt{N}\, p_i\, [\bar{Y}_{\hat{\xi}_{p_i}} - E(Y_j | Y_j \leq \xi_{p_i})]$$

$$\doteq \sqrt{N}\, [N^{-1}\, \Sigma_{Y_j \leq \hat{\xi}_{p_i}} Y_j - \tau_i]. \qquad (A3)$$

Then consider the first term in (A3):

$$\Sigma_{Y_j \leq \hat{\xi}_{p_i}} Y_j = \Sigma_1^N I_j Y_j + \Sigma_{Y_j \varepsilon (\xi_{p_i},\, \hat{\xi}_{p_i})} Y_j$$

$$\doteq \Sigma_1^N I_j Y_j + \xi_{p_i} R + 0_p(1) \qquad (A4)$$

where it is assumed for convenience that $\xi_{p_i} < \hat{\xi}_{p_i}$, and where R represents the (signed) number of observations between $\hat{\xi}_{p_i}$ and ξ_{p_i}. Since the number of observations in a small interval of length Δ about ξ_{p_i} is approximately $Nf(\xi_{p_i})\Delta$, and since the (signed) length of the interval between $\hat{\xi}_{p_i}$ and ξ_{p_i} is approximately

$$N^{-\frac{1}{2}}[p_i(1-p_i)]^{\frac{1}{2}}\, \varepsilon / f(\xi_{p_i}) \text{ from (A1)},$$

$$R \doteq N^{\frac{1}{2}}[p_i(1-p_i)]^{\frac{1}{2}} \varepsilon. \qquad (A5)$$

Thus, from (A4) and (A5),

$$S_i = N^{-\frac{1}{2}}\, \Sigma_1^N\, (I_j Y_j - \tau_i) + \xi_{p_i}\, [p_i(1-p_i)]^{\frac{1}{2}} \varepsilon + 0_p(1),$$

where the first term is asymptotically normal with mean zero by the Central Limit Theorem, and the second has also been shown to be asymptotically normal with mean zero in Theorem 1 of the text. Consequently, S_i is also asymptotically normal with mean zero, and $p_i \bar{Y}_{\hat{\xi}_{p_i}}$ is asymptotically normal with mean $p_i\, E(Y_j | Y_j \leq \xi_{p_i}) = \tau_i$.

Now, by the argument in Section III.1, the limiting distribution of a continuous function of asymptotically normal random variables is also asymptotically normal. In particular, consider the ratio $p_i \bar{Y}_{\hat{\xi}_{p_i}} / \bar{Y}$ both of whose arguments have been shown to be asymptotically normal with means τ_i and μ respectively. Then it follows that

$$\sqrt{N}\left[p_i \frac{\bar{Y}_{\hat{\xi}_{p_i}}}{\bar{Y}} - \frac{\tau_i}{\mu}\right] = \sqrt{N}(\hat{\Phi}_i - \Phi_i)$$

is also asymptotically normal with mean zero and a constant variance for $i = 1, \ldots, K$. That is, $\sqrt{N}(\hat{\Phi}_i - \Phi_i)$ and $\sqrt{N}(\Phi(\hat{\xi}_{p_i}) - \Phi_i)$ have the same probability limit of zero, so that the feasible estimator $\hat{\Phi}_i$ and the infeasible estimator $\Phi(\hat{\xi}_{p_i})$ are asymptotically equivalently distributed for all $i = 1, \ldots, K$ [27, p. 101].

<u>References</u>

[1] Aigner, D. J., and A. S. Goldberger: "Estimation of Pareto's Law from Grouped Observations" *Journal of the American Statistical Association*, 65 (1970), 712-723.

[2] Atkinson, A. B.: "On the Measurement of Inequality", *Journal of Economic Theory*, 2 (1970), 244-263.

[3] Bahadur, R. R.: "A Note on Quantiles in Large Samples", *Annals of Mathematical Statistics*, 37 (1966), 577-580.

[4] Beach, C. M.: "Cyclical Sensitivity of Aggregate Income Inequality", *Review of Economics and Statistics*, 59 (1977), 56-66.

[5] __________: "Inference with the Relative Mean Income Curve and Associated Inequality Measures", Discussion Paper, Dept. of Economics, Queen's University (1979), forthcoming.

[6] __________, with the assistance of D. E. Card and F. Flatters: *Distribution of Income and Wealth: Theory and Evidence for Ontario*. Toronto: Ontario Economic Council, 1980.

[7] Bradley, J. V.: *Distribution-Free Statistical Tests*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

[8] Chernoff, H., J. L. Gastwirth, and M. V. Johns, Jr.: "Asymptotic Distribution of Linear Combinations of Functions of Order Statistics with Applications to Estimation", *Annals of Mathematical Statistics*, 38 (1967), 52-72.

[9] Danziger, S. and M. K. Taussig: "The Income Unit and the Anatomy of Income Distribution", Institute for Research on Poverty Discussion Paper No. 516-78, University of Wisconsin - Madison, 1978.

[10] Donaldson, D. and J. A. Weymark: "A Single-Parameter Generalization of the Gini Indices of Inequality", Discussion Paper 79-11, Dept. of Economics, University of British Columbia, 1979.

[11] Dorfman, R.: "A Formula for the Gini Coefficient", *Review of Economics and Statistics,* 61 (1979), 146-149.

[12] Elteto, O., and E. Frigyes: "New Income Inequality Measures as Efficient Tools for Causal Analysis and Planning", *Econometrica,* 36 (1968), 383-396.

[13] Fair, R.C.: "The Optimal Distribution of Income", *Quarterly Journal of Economics,* 85 (1971), 551-579.

[14] Gastwirth, J. L.: "A General Definition of the Lorenz Curve", *Econometrica,* 39 (1971), 1037-1039.

[15] __________: "The Estimation of the Lorenz Curve and Gini Index", *Review of Economics and Statistics,* 54 (1972), 306-316.

[16] __________: "Large Sample Theory of Some Measures of Income Inequality", *Econometrica,* 42 (1974), 191-196.

[17] Gastwirth, J. L., and M. Glauberman: "The Interpolation of the Lorenz Curve and Gini Index from Grouped Data", *Econometrica,* 44 (1976), 479-484.

[18] Kakwani, N. C.: "A Note on the Efficient Estimation of the New Measures of Income Inequality", *Econometrica,* 42 (1974), 597-600.

[19] Kakwani, N. C. and N. Podder: "On the Estimation of Lorenz Curves from Grouped Observations", *International Economic Review,* 14 (1973) 278-292.

[20] __________: "Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations", *Econometrica*, 44 (1976), 137-148.

[21] Kendall, M. G. and A. Stuart: *The Advanced Theory of Statistics*, 3rd Ed., Vol. I. London: Charles Griffen & Co., 1969.

[22] Kloek, T., and H. K. van Dijk: "Further Results on Efficient Estimation of Income Distribution Parameters", *Economie Appliquée*, 30 (1977) 1-21.

[23] __________: "Efficient Estimation of Income Distribution Parameters", *Journal of Econometrics*, 8 (1978), 61-74.

[24] Levine, D. B., and N. M. Singer: "The Mathematical Relation Between the Income Density Function and the Measurement of Income Inequality", *Econometrica*, 38 (1970), 324-330.

[25] Martić, L.: "A Geometrical Note on New Income Inequality Measures", *Econometrica*, 38 (1970), 936-937.

[26] Mosteller, F.: "On Some Useful Inefficient Statistics", *Annals of Mathematical Statistics*, 17 (1946), 377-407.

[27] Rao, C. R.: *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons, 1965.

[28] Sarhan, A. E., and B. G. Greenberg, ed.: *Contributions to Order Statistics*. N.Y.: John Wiley & Sons, 1962.

[29] Scheffé, H.: *The Analysis of Variance*. N.Y.: John Wiley & Sons, 1959.

[30] Seber, G. A. F.: *Linear Regression Analysis*. N.Y.: John Wiley & Sons, 1977.

[31] Sen, A.: *On Economic Inequality*. Oxford: Clarendon Press, 1973.

[32] Thurow, L. C.: "Analyzing the American Income Distribution", *American Economic Review*, 60 (1970), 261-270.

[33] Wilks, S. S.: *Mathematical Statistics*. N.Y.: John Wiley & Sons, 1962.

[34] Wold, H.: "A Study on the Mean Difference, Concentration Curves, and Concentration Ratio", *Metron*, 12 (1935), 39-58.