Queen's Economics Department Working Paper No. 1484

Better Off or More Apart? Empirically Testing Welfare and Inequality Dominance Criteria

Charles Beach

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

3-2022

Better Off or More Apart? Empirically Testing Welfare and Inequality Dominance Criteria

by

Charles M. Beach

Department of Economics

Queen's University

Kingston, ON, Canada

March 2022

Abstract

This paper provides the tools and procedures for empirically implementing several dominance criteria for social welfare comparisons and broad income inequality comparisons. Dominance criteria are expressed in terms of vectors of quantile ordinates based on income shares or quantile means. Statistical properties of these sample ordinates are established that allow a framework for statistical inference on these vectors. And practical empirical criteria are forwarded for using formal statistical inference tests to reach conclusions about ranking social welfare and inequality between distributions. Examples include rank dominance, generalized Lorenz dominance, dominance with crossing Lorenz curves, and distributional distance dominance between income groups.

1. <u>Introduction</u>

In a time when rising wages coincide with increased inequality it is reasonable to ask whether people as a whole are better or worse off? But to answer such a question, one needs an empirical criterion for "better off" or for economic well-being. Similarly, recent focus on issues such as "common prosperity" (The Economist, 2021a,b), "equitable growth" (Drummond, 2021), "fairness in growth" (Lohr, 2022), and "quality of life" (Department of Finance Canada, 2021) further enhances the need for such an empirically implementable criterion. Again, will government policies focussed on helping the so-called Middle Class (Wright, 2021) or reducing income inequality benefit overall economic well-being? An empirically implementable criterion could also allow one to quantify how economic well-being has varied over time for a given economy or region or to formally test for statistically significant differences in the criterion between countries such as Canada and the United States. The present paper offers an empirical approach to evaluating changes or differences in economic well-being in an easily implementable framework of statistical inference based on disaggregative distributional statistics. This approach could also be part of a burgeoning interest in the broad endeavour of distributional National Accounts (see, for example, Alvaredo et al., 2018, 2020, and Zucman et al., 2018).

More specifically, the paper applies the technical developments in Beach (2021a,b) which establish a statistical inference framework for a toolbox of disaggregative inequality measures in order to forward an empirically implementable set of rules or a "practical empirical criterion" (PEC) for establishing when economic welfare has changed statistically significantly. The paper applies this inference approach and PEC to several dominance rules (including rank dominance, Lorenz dominance, and generalized Lorenz dominance) provided in the theoretical welfare literature. In so doing, it extends the set of toolbox measures of income inequality, develops a distributional distance function concept, and provides explicit formulas for distribution-free standard errors of Lorenz curve ordinates and related measures.

The main contributions of the paper are thus: (i) showing how to empirically implement several welfare dominance criteria in terms of vectors of quantile-based distributional statistics; (ii) presenting formulas for (distribution-free) asymptotic variances and covariances – and hence standard errors – of these distributional statistics; (iii) forwarding a practical empirical criterion, or PEC, for establishing the stochastic dominance of one vector of distributional statistics over another; and (iv) developing a new distributional distance vector construct that can also be used in empirical dominance analysis. The proposed empirical inference framework can thus be used to help analyze, say, the distributional and economic welfare consequences of macroeconomic events or policies, or of demographic changes such as the oncoming avalanche of Baby Boomer retirements (perhaps hastened by the covid pandemic), or more generally to empirically test for first- and second-order stochastic dominance situations in a formal statistical inference framework.

The paper is organized as follows. The next section sets out the quantile function approach developed in previous work by the author (Beach 2021a,b) and that serves as the basis for the statistical inference framework of the present analysis. Asymptotic variance and standard error results are presented for quantile-based estimates of sets of quantile means and income shares – such as typically provided by Statistics Canada or the United States Bureau of the Census – used to distributionally characterize a distribution of income. Section 3 outlines the normative perspective to evaluating changes in social welfare and inequality. Then Section 4 applies the above empirical framework to testing for rank dominance between distributions.

Since each distribution is characterized by a vector of disaggregative distributional statistics, the comparison of two vectors involves setting out a practical empirical criterion (or PEC) for the statistical ranking of vectors of quantile mean random variables. Section 5 and 6 then apply this empirical framework to Lorenz dominance and generalized Lorenz dominance between distributions with appropriate PECs forwarded for each. Section 7 shows how one can decompose social welfare into efficiency and equity contribution components and develops a PEC for the equity component as well. Section 8 and 9 consider how to address inequality dominance when the Lorenz curves of two distributions cross one or more times. And Section 10 applies the PEC approach to comparing vectors of "distributional distance" characterizing two income distributions. Then Section 11 concludes with some implications of the paper's analysis.

2. Quantile Function Approach Basic Results

2.1 Empirical Framework of Analysis

Empirical measures of economic well-being and inequality are built up from disaggregative statistics on percentile mean income levels and percentile income shares. Percentile statistics are those that are expressed in terms of given percentage groups of the ranked or ordered observations in a microdata sample. In the case of income distribution statistics, the data observations in a sample are ordered by income from the lowest income observation to the highest income observation. The ordered observations are then divided into non-overlapping income groups, say, in terms of ten deciles (or generically referred to as quantile income groups or simply quantiles). So the first decile group consists of those observations with the 10 percent lowest income levels, the second decile group consists of the next 10 percent lowest income recipients, and so on up to the top or tenth decile income group which includes those 10 percent of income recipients with the highest income levels in the sample. The standard Lorenz curve of (cumulated) income shares, for example, is based around such percentile groups, and quantile mean income levels can also be calculated for each of the percentile groups.

The key feature of such percentile statistics is that the relative sizes of the percentile groups are *given* percentages of the sample or distribution. Quantile means and quantile income share are two examples of what can be referred to as toolbox measures (Beach, 2021a) of characterizing the disaggregative structure of an income distribution. Both Statistics Canada and the U.S. Bureau of the Census publish annual series on both decile (and quintile) income shares and decile/quintile mean income levels. The empirical analysis of this paper focuses on quantile statistics as a way to characterize the detailed structure of distribution because, as shown in sections 2.3 and 2.4 below, statistical inference calculations based on them turn out to have an especially straightforward and convenient form.

2.2 The Quantile Function Approach to Statistical Inference

The income share and quantile mean statistics are calculated from sample survey data, and hence can be viewed as sample estimates of their corresponding features in the (unobserved) overall underlying income distribution. They can thus be viewed as random variables with associated sampling distributions. What we want to do is to figure out what one can say about these sampling distributions, so that one can undertake formal statistical inference on these estimated measures. The so-called quantile function approach is a way to address this problem (Beach, 2021a,b).

Consider first some formal concepts and notation. Suppose the distribution of income *Y*, is divided into *K* ordered income groups, so that $K = 10$ in the case of deciles and $K = 5$ for quintiles. Let the dividing proportions of recipients be $p_1 < p_2 < \cdots < p_{K-1}$ (with $p_0 = 0$ and $p_K = 1.0$).[1] Then in terms of the underlying (population) density of income recipients, the mean income of the i'th quantile is given by

$$\mu_i = \int_{\xi_{i-1}}^{\xi_i} y\, f(y) dy \,/\, \int_{\xi_{i-1}}^{\xi_i} f(y) dy \qquad \text{for } i = 1, \ldots, K \qquad (1)$$

where $f(\bullet)$ is the underlying population density function and the ξ_i's are the cut-off income levels corresponding to the proportions $p_1, p_2, \ldots, p_{K-1}$ (with $\xi_0 = 0$). Since the income group proportions are given for percentile statistics, the denominator in (1) is given by

$D_i = p_i - p_{i-1}$, so that

$$\mu_i = \left(\frac{1}{D_i}\right) \int_{\xi_{i-1}}^{\xi_i} y\, f(y) dy \; . \qquad (2)$$

This integral expression – what we'll refer to as a quantile function – links the quantile mean μ_i to the quantile cut-offs ξ_i , ξ_{i-1} . It turns out that a powerful theorem by C.R. Rao (1965) says that, if we know the asymptotic distribution of the sample estimates $\hat{\xi}_i$ and $\hat{\xi}_{i-1}$ as asymptotically joint normal and if, in the population, μ_i can be expressed as a continuous and differentiable function of ξ_i and ξ_{i-1} , then the sample estimate $\hat{\mu}_i$ will also be asymptotically normally distributed with (asymptotic) mean μ_i and (asymptotic) variance that can be easily calculated in terms of first derivatives of expression (2). We will refer to this as Rao's linkage theorem. Since the asymptotic distribution of the sample cut-offs $\hat{\xi}_i$'s has long been well established, this theorem provides the basis of the quantile function approach (or QFA) used in

[1] We assume in what follows that the data samples used are random samples. If the survey records are indeed weighted, the formulas can be readily adjusted by replacing sums of observations by sums of the sample weighted observations.

Beach (2021a,b) and the present paper. The basic idea is to express the various percentile distributional measures in terms of integral functions of the income cut-offs (the ξ_i's) and then invoke Rao's linkage theorem to establish asymptotic normality and expressions for the sample measures' asymptotic variances. Standard errors, then, are simply obtained from these estimated (asymptotic) variances rescaled by the size of the estimation sample:

$$S.E.(\hat{\mu}_i) = \left[\frac{Asy\hat{.}var(\hat{\mu}_i)}{N}\right]^{1/2}$$

where *N* is the sample size of the estimation sample.

Now, in general one would expect the (asymptotic) variances to depend on the specific functional form of the underlying income distribution's density $f(\bullet)$. Certainly the (asymptotic) variance-covariance structure of the $\hat{\xi}_i$'s does. But – as will be shown in the next two subsections – perhaps surprisingly, the resulting (asymptotic) variances and standard errors of the percentile-based distributional measures are a special case that do *not* depend upon the specific functional form of $f(\bullet)$. In this sense, they are said to be distribution-free, and hence very straightforward to calculate. Taking a quantile function approach thus allows one to avoid having to estimate assumed underlying population density function forms (such as the lognormal in Beach, 2021a) or to undertake burdensome bootstrapping estimation techniques for density ordinate evaluation (as in Davidson, 2018).

2.3 Application of the QFA to Quantile Means

The starting point is to establish the asymptotic distribution and its variance-covariance structure for the full set of sample quantile income cut-off levels. Suppose that the income distribution is divided into *K* ordered income groups corresponding to the cumulative proportions $0 < p_1 < p_2 < \cdots < p_K = 1$ and the quantile cut-offs $\xi_1, \xi_2, \ldots \xi_{K-1}$. Let $\hat{\xi} =$

$(\hat{\xi}_1, \hat{\xi}_2, \ldots \hat{\xi}_{K-1})'$ be a vector of *K-1* sample quantile cut-offs[2] from a random sample of size N drawn from a continuous population density $f(\bullet)$ such that the $\hat{\xi}_i$'s are uniquely defined and $f_i \equiv f(\xi_i) > 0$ for all *i = 1, ..., K-1*. Then it can be proved (see, for example, Wilks (1962), p. 273, or Kendall and Stuart (1969, pp. 237-239)) that the vector $\sqrt{N}\,(\hat{\xi} - \xi)$ converges in distribution to a *(K-1)*-variate normal distribution with mean zero and variance-covariance matrix **Λ** where

$$\Lambda = \begin{bmatrix} \frac{p_1(1-p_1)}{f_1^2} & \cdots & \frac{p_1(1-p_{K-1})}{f_1 f_{K-1}} \\ \vdots & & \vdots \\ \frac{p_1(1-p_{K-1})}{f_1 f_{K-1}} & \cdots & \frac{p_{K-1}(1-p_{K-1})}{f_{K-1}^2} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1,K-1} \\ \vdots & & \vdots \\ \lambda_{1,K-1} & \cdots & \lambda_{K-1,K-1} \end{bmatrix} = [\lambda_{ij}] \;. \qquad (3)$$

Note how the (asymptotic) variances and covariances explicitly depend on the specific functional form of $f(\bullet)$ in the denominators of the λ_{ij}'s.

Then applying a multivariate version of Rao's linkage theorem (Rao, 1965, p. 388), consider the full set of *K* sample quantile means $\hat{m} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K)'$ corresponding to the vector of population quantile means $m = (\mu_1, \mu_2, \ldots, \mu_K)'$ where μ_i is defined in eq. (2). In the case of deciles, *K = 10* and $D_i = 0.10$. Then according to Rao's theorem for continuous differentiable functions, the vector $\hat{m}$ is asymptotically joint normally distributed in that $\sqrt{N}(\hat{m} - m)$ converges in distribution to a joint normal with *KxK* (asymptotic) variance-covariance matrix *V* where

$$Asy.var(\hat{m}) \equiv V = G\,\Lambda\,G' \qquad (4a)$$

and the *K x (K-1)* matrix *G* is

[2] To estimate the sample quantile cut-offs, order the sample of *N* observations by income level. Then, in the case of deciles, $\hat{\xi}_i$ is that income level such that $p_i N$ observations lie below it and the rest above. If there is no single observation meeting this condition, simply take the average of the two adjacent observations (below and above) that are closest.

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1,K-1} \\ \vdots & & \vdots \\ g_{K,1} & \cdots & g_{K,K-1} \end{bmatrix} = \left[g_{ij}\right]$$

$$= \left[\frac{\partial \mu_i}{\partial \xi_j}\right] \qquad \text{with } i = 1, \ldots, K \text{ rows}$$

and $j = 1, \ldots, K\text{-}1$ columns. (4b)

For convenience, rewrite eq. (2) as

$$\mu_i = \left(\frac{1}{D_i}\right) \bullet N_i(\xi_i, \xi_{i-1}) \qquad \text{for } i = 1, \ldots, K,$$

where N_i is an explicit function of ξ_i and ξ_{i-1} in the numerator of the expression for μ_i .

In deriving the components of $\left[g_{ij}\right]$, let us illustrate with the case of decile income groups. Then it can be worked out that

$$g_{11} = \frac{\partial \mu_1}{\partial \xi_1} = 10 \frac{\partial N_1}{\partial \xi_1} = 10\, \xi_1 \bullet f(\xi_1)$$

$$g_{1j} = \frac{\partial \mu_1}{\partial \xi_j} = 10 \frac{\partial N_1}{\partial \xi_j} = 0 \qquad \text{for } j = 2, \ldots, K\text{-}1 .$$

$$g_{21} = \frac{\partial \mu_2}{\partial \xi_1} = 10 \frac{\partial N_2}{\partial \xi_1} = 10\,(-\xi_1) \bullet f(\xi_1)$$

$$g_{22} = \frac{\partial \mu_2}{\partial \xi_2} = 10 \frac{\partial N_2}{\partial \xi_2} = 10\, \xi_2 \bullet f(\xi_2)$$

$$g_{2j} = \frac{\partial \mu_2}{\partial \xi_j} = 10 \frac{\partial N_2}{\partial \xi_j} = 0 \qquad \text{for } j = 3, \ldots, K\text{-}1 .$$

$$g_{Kj} = \frac{\partial \mu_K}{\partial \xi_j} = 10 \frac{\partial N_K}{\partial \xi_j} = 0 \qquad \text{for } j = 1, \ldots, K\text{-}2 .$$

$$g_{K,K-1} = \frac{\partial \mu_K}{\partial \xi_{K-1}} = 10 \frac{\partial N_K}{\partial \xi_{K-1}} = 10\,(-\xi_{K-1}) \bullet f(\xi_{K-1}).$$

As a result, the G matrix is the banded diagonal-type matrix:

$$G = [\begin{matrix} 10\,\xi_1 \bullet f(\xi_1) & 0 & 0 & 0 & \cdots \\ -10\,\xi_1 \bullet f(\xi_1) & 10\,\xi_2 \bullet f(\xi_2) & 0 & 0 & \cdots \\ 0 & -10\,\xi_2 \bullet f(\xi_2) & 10\,\xi_3 \bullet f(\xi_3) & 0 & \cdots \\ \vdots & 0 & -10\,\xi_3 \bullet f(\xi_3) & 10\,\xi_4 \bullet f(\xi_4) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots \end{matrix}$$

$$
\begin{matrix}
\cdots & 0 & 0 \\
\cdots & 0 & 0 \\
 & \vdots & \vdots \\
\cdots & 0 & 0 \\
\cdots & -10\,\xi_8 \bullet f(\xi_8) & 10\,\xi_9 \bullet f(\xi_9) \\
\cdots & 0 & -10\,\xi_9 \bullet f(\xi_9)
\end{matrix} \quad] \, . \qquad (5)
$$

The (asymptotic) variances, then, are gotten by multiplying the corresponding row of *G* and column of G' (i.e., row of *G*) by the appropriate diagonal element of the variance-covariance matrix . So

$$
\begin{aligned}
Asy.var(\hat{\mu}_1) &= G(row\ 1) \bullet \Lambda \bullet\ G(row\ 1)' \\
&= (10)^2\, \xi_1^2\ \bullet f(\xi_1)^2\ \bullet\ \left[\frac{p_1(1-p_1)}{f(\xi_1)^2}\right] \\
&= (10)^2\, p_1(1-p_1)\, \xi_1^2\ . \qquad (6a)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
Asy.var(\hat{\mu}_{10}) &= G(row\ 10) \bullet \Lambda \bullet\ G(row\ 10)' \\
&= (10)^2\, \xi_9^2\ \bullet f(\xi_9)^2\ \bullet\ \left[\frac{p_9(1-p_9)}{f(\xi_9)^2}\right] \\
&= (10)^2\, p_9(1-p_9)\, \xi_9^2\ . \qquad (6b)
\end{aligned}
$$

And for *i = 2, …, 9* ,

$$
\begin{aligned}
Asy.var(\hat{\mu}_i) &= G(row\ i) \bullet \Lambda \bullet\ G(row\ i)' \\
&= (10)^2\, [p_{i-1}(1-p_{i-1})\xi_{i-1}^2 + p_i(1-p_i)\, \xi_i^2 \\
&\qquad - 2\, p_{i-1}(1-p_i)\xi_{i-1}\, \xi_i]\ . \qquad (6c)
\end{aligned}
$$

More generally, then,

$$
Asy.var(\hat{\mu}_1) = \left(\frac{1}{D_1}\right)^2\ p_1(1-p_1)\, \xi_1^2\ ,
$$

$$
Asy.var(\hat{\mu}_K) = \left(\frac{1}{D_K}\right)^2\ p_{K-1}(1-p_{K-1})\, \xi_{K-1}^2\ ;
$$

and for *i = 2, …, K-1* ,

$$Asy.var(\hat{\mu}_i) = \left(\frac{1}{D_{i-1}}\right)^2 p_{i-1}(1-p_{i-1})\xi_{i-1}^2 + \left(\frac{1}{D_i}\right)^2 p_i(1-p_i)\xi_i^2$$

$$- 2\left(\frac{1}{D_{i-1}}\right)\left(\frac{1}{D_i}\right) p_{i-1}(1-p_i)\xi_{i-1}\,\xi_i \,. \tag{7}$$

If the proportional size of each income group is the same, so that $D_i = \left(\frac{1}{K}\right)$ for all *i = 1, ..., K,* then

$$Asy.var(\hat{\mu}_1) = K^2\, p_1(1-p_1)\,\xi_1^2 \tag{8a}$$

$$Asy.var(\hat{\mu}_K) = K^2\, p_{K-1}(1-p_{K-1})\,\xi_{K-1}^2 \tag{8b}$$

and $$Asy.var(\hat{\mu}_i) = K^2\,[p_{i-1}(1-p_{i-1})\xi_{i-1}^2 + p_i(1-p_i)\xi_i^2$$

$$- 2\, p_{i-1}(1-p_i)\xi_{i-1}\,\xi_i] \tag{8c}$$

for *i = 2, ..., K-1.*

These results on the (asymptotic) variances, then, are sufficient to determine the standard errors of the quantile mean estimates. Since the formulas in eqs. (6)-(8) involve unknown population parameters, one obtains *estimated* (asymptotic) variances by replacing all the unknown parameters by their consistent estimates. So, for example, in (6a),

$$\widehat{Asy.var}(\hat{\mu}_1) = (10)^2\, p_1(1-p_1)\,\hat{\xi}_1^2$$

where ξ_1 is replaced by its standard sample estimate. Rao (1965, p. 355) has also shown that if $f(\bullet)$ is strictly positive, then the $\hat{\xi}_i$'s are indeed (strongly) consistent. The resulting standard error for $\hat{\mu}_1$ is then gotten by adjusting for the sample size of the estimation sample:

$$S.E.(\hat{\mu}_1) = \left[\frac{\widehat{Asy.var}(\hat{\mu}_1)}{N}\right]^{1/2}.$$

Or more generally,

$$S.E.(\hat{\mu}_i) = \left[\frac{\widehat{Asy.var}(\hat{\mu}_i)}{N}\right]^{1/2} \tag{9}$$

for all *i = 1, ..., K* .

Note as well that the asymptotic variances and standard errors of the quantile means for given percentile groups are distribution-free. This is because of the way that the $f(\xi_i)$ terms all cancel out in the derivation in the case of percentile measures. The formulas in eqs. (6)-(9) are thus very straightforward and easy to calculate.

One can apply these results to look at *differences* in individual quantile means between different population groups – such as quantile mean earnings differences between male and female workers in the labour market – and at *changes* in separate quantile means between time periods. So long as the estimates being compared are from independent samples, the variance of the difference in sample estimates is simply the sum of the separate variances, and the standard error of the difference is given by

$$S.E.\left(\hat{\mu}_i^b - \hat{\mu}_i^a\right) = \left[\frac{Asy\hat{\cdot}var\left(\hat{\mu}_i^b\right)}{N^b} + \frac{Asy\hat{\cdot}var\left(\hat{\mu}_i^a\right)}{N^a}\right]^{1/2} \tag{10}$$

where superscripts *a* and *b* refer to the two separate sample estimates. A quantile analysis thus allows for potentially quite detailed disaggregative examination of differences between distributions. And the range of toolbox measures available furthers the perspective and flexibility of such examinations.

Indeed, one could express these differences in relative or percentage terms – or what The Economist (2021c, p. 24) refers to as Piketty lines of different growth rates of quantile means across the different regions of the income distribution. In this case, it is shown in Beach (2021b, p. 15) that, if

$$\hat{q}_i = \left(\hat{\mu}_i^b - \hat{\mu}_i^a\right) / \hat{\mu}_i^a = \left(\frac{\hat{\mu}_i^b}{\hat{\mu}_i^a}\right) - 1\,,$$

then approximately

$$V\hat{a}r(\hat{q}_i) = \left(\frac{-\hat{\mu}_i^b}{\left(\hat{\mu}_i^a\right)^2}\right)^2 \bullet\ V\hat{a}r(\mu_i^a) + \left(\frac{1}{\hat{\mu}_i^a}\right)^2 \bullet V\hat{a}r\left(\hat{\mu}_i^b\right)$$

$$= \left(\frac{-\hat{\mu}_i^b}{\left(\hat{\mu}_i^a\right)^2}\right)^2 \bullet \left[\frac{Asy\hat{.}var\left(\hat{\mu}_i^a\right)}{N^a}\right] + \left(\frac{1}{\hat{\mu}_i^a}\right)^2 \bullet \left[\frac{Asy.var\left(\hat{\mu}_i^b\right)}{N^b}\right] \quad (11)$$

and again

$$S.E.(\hat{q}_i) = [V\hat{a}r(\hat{q}_i)]^{1/2}.$$

Again, the standard error estimates are distribution-free.

It would thus be helpful to users of official decile and quintile mean statistics from government statistical agencies if these agencies provided the actual sample sizes (i.e., the *N* in the denominator of (9)) that their survey estimates are based on, and not just the overall survey sample size.

2.4 Application of the QFA to Incomes Shares

The income share of the i'th income group can be expressed as

$$IS_i \equiv \int_{R_i} \left(\frac{1}{\mu}\right) yf(y)dy \qquad \text{for } i = 1, \ldots, K, \quad (12a)$$

with integration over the region R_i running from ξ_{i-1} to ξ_i , and μ is the mean of the overall (population) distribution of income. The integral in (12a) can, for future notational convenience, be written as

$$IS_i = N_i(\xi_{i-1}, \xi_i, \mu) = \int_{\xi_{i-1}}^{\xi_i} \left(\frac{1}{\mu}\right) yf(y)dy\,. \quad (12b)$$

It can be seen that estimates of IS_i or N_i involve estimates of two sets of parameters – the range of integration cut-off ξ_{i-1} and ξ_i and the overall population mean μ . To take account of this, we make use of a useful paper by Lin, Wu and Ahmad (1980) (henceforth LWA). LWA establish that, under general regularity conditions, $\hat{\xi}_i$, $\hat{\xi}_{i-1}$, and $\hat{\mu}$ are asymptotically joint normally distributed with (asymptotic) variance-covariance matrix

$$\Sigma = \left[\sigma_{ij}\right] \quad (13)$$

where $\sigma_{11} = \frac{p_{i-1}(1-p_{i-1})}{[f(\xi_{i-1})]^2}$, $\sigma_{22} = \frac{p_i(1-p_i)}{[f(\xi_i)]^2}$, $\sigma_{33} = \sigma^2$

$$\sigma_{12} = \frac{p_{i-1}(1-p_i)}{f(\xi_{i-1})f(\xi_i)} = \sigma_{21}$$

$$\sigma_{13} = \frac{\xi_{i-1} - \mu(1-p_{i-1})}{f(\xi_{i-1})} = \sigma_{31}$$

and $\sigma_{23} = \frac{\xi_i - \mu(1-p_i)}{f(\xi_i)} = \sigma_{32}$,

where σ^2 is the variance of the overall (population) distribution of income.

One can now combine this set of LWA result with Rao's linkage theorem. So, if $\hat{\xi}_{i-1}, \hat{\xi}_i$ and $\hat{\mu}$ are asymptotically joint normal with (asymptotic) variance-covariance matrix Σ above, then the (asymptotic) variance of $\widehat{IS}_i$ is given by

$$Asy.var(\widehat{IS}_i) = G' \, \Sigma \, G \quad (14)$$

where

$$G = \left[\frac{\partial N_i}{\partial \xi_{i-1}}, \frac{\partial N_i}{\partial \xi_i}, \frac{\partial N_i}{\partial \mu}\right]' = [g_1, g_2, g_3]' .$$

So in the case of *i = 1*:

$$g_1 = 0$$

$$g_2 = \left(\frac{1}{\mu}\right) \xi_1 f(\xi_1)$$

$$g_3 = \frac{-N_1}{\mu} = \frac{-IS_1}{\mu} ,$$

and

$$Asy.var(I\hat{S}_1) = g_2^2 \, \sigma_{22} + g_3^2 \, \sigma_{33} + 2g_2 g_3 \sigma_{23}$$

$$= \left(\frac{\xi_1}{\mu}\right)^2 p_1(1-p_1) + \left(\frac{IS_1}{\mu}\right)^2 \sigma^2 - 2 \left(\frac{\xi_1}{\mu}\right)\left(\frac{IS_1}{\mu}\right) [\xi_1 - \mu(1-p_1)] . \quad (15)$$

In the case of *i = 10*:

$$g_1 = -\left(\frac{1}{\mu}\right) \xi_9 \bullet f(\xi_9)$$

$$g_2 = 0$$

$$g_3 = \frac{-N_{10}}{\mu} = \frac{-IS_{10}}{\mu},$$

so

$$Asy.var\left(I\hat{S}_{10}\right) = g_1^2\, \sigma_{11} + g_3^2\, \sigma_{33} + 2 g_1 g_3 \sigma_{13}$$

$$= \left(\frac{\xi_9}{\mu}\right)^2 p_9(1-p_9) + \left(\frac{IS_{10}}{\mu}\right)^2 \sigma^2 + 2\left(\frac{\xi_9}{\mu}\right)\left(\frac{IS_{10}}{\mu}\right)[\xi_9 - \mu(1-p_9)]\,. \qquad (16)$$

And in the case of $i = 2, \ldots, 9$:

$$g_1 = -\left(\frac{1}{\mu}\right) \xi_{i-1} \bullet f(\xi_{i-1})$$

$$g_2 = \left(\frac{1}{\mu}\right) \xi_i \bullet f(\xi_i)$$

and $$g_3 = -\left(\frac{1}{\mu}\right) IS_i\,.$$

Therefore, $Asy.var\left(I\hat{S}_i\right) = G'\, \Sigma\, G$

$$= \left(\frac{\xi_{i-1}}{\mu}\right)^2 p_{i-1}(1-p_{i-1}) + \left(\frac{\xi_i}{\mu}\right)^2 p_i(1-p_i) + \left(\frac{IS_i}{\mu}\right)^2 \sigma^2$$

$$- 2\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{\xi_i}{\mu}\right) p_{i-1}(1-p_i)$$

$$+ 2\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{IS_i}{\mu}\right)[\xi_{i-1} - \mu(1-p_{i-1})] \qquad (17)$$

$$- 2\left(\frac{\xi_i}{\mu}\right)\left(\frac{IS_i}{\mu}\right)[\xi_i - \mu(1-p_i)]\,.$$

The standard error of the i'th quantile income share is thus given by

$$S.E.\left(\widehat{IS}_i\right) = \left[\frac{Asy\hat{.}var(\widehat{IS}_i)}{N}\right]^{1/2}.$$

Note, incidentally, that just as IS_i is a ratio and hence units-free, so also is each term of its (asymptotic) variance and hence its standard error. The effect of randomness operating through $\hat{\mu}$

operates through the third term (corresponding to the simple variance of $\hat{\mu}$) and the last two terms (corresponding to $\hat{\mu}$'s covariance with $\hat{\xi}_{i-1}$ and $\hat{\xi}_i$, respectively).

And, again, the standard error formulas for income shares are also distribution-free, and conventional statistical inference can be undertaken in straightforward fashion.

Note that the formulas in equations (15)-(17) include σ^2, so that implementation of these formulas requires an estimate of the variance (or standard deviation) for the sample used to calculate the terms in (15)-(17). So again it would be helpful to users of quantile income share statistics from official statistical agencies if these agencies provided the estimated standard deviations for the actual samples that the survey estimates are based on, and not just the estimated sample means.

Quantile means and income shares serve as the basis for operationally implementing the evaluation of changes in social welfare and income inequality.

3. A Normative Perspective for Evaluating Changes in Social Welfare and Inequality

The traditional way of measuring income inequality in an income distribution is in terms of some summary or aggregate measure of inequality such as the Gini coefficient (G), coefficient of variation (C), relative mean (absolute) deviation (M), or the standard deviation of the logs of income (L). But such measures are subject to two basic criticisms. First is the aggregation problem: various summary measures aggregate income differences in different ways, so that different measures can give different results when comparing two distributions. One way (partially) to address this is to identify several desirable properties we may want such summary measures to satisfy. These could include, for example:

i) Symmetry (or Anonymity) – An inequality measure depends only on incomes in a distribution and not on who has which incomes;

ii) Mean Independence – An inequality measure is invariant to proportional changes (e.g., doubling) of all incomes (i.e., it is a relative measure of inequality);

iii) Population Homogeneity – An inequality measure is invariant to replication of the population (e.g., doubling the number of persons in the distribution while keeping the shape of the distribution the same);

iv) Principle of Transfers – Any transfer of \$x from a richer person to a poorer person so that $y_i + x < y_j - x$ if initially $y_i < y_j$ should reduce inequality;

v) Transfer Sensitivity – A transfer of \$x such as envisioned in (iv) should reduce inequality more if it occurs among a lower-income pair of individuals than if it occurs among a higher-income pair of individuals. This is obviously a stronger form of the Principle of Transfers.

It turns out that, (i), (ii) and (iii) are satisfied by all the above four inequality measures, but (iv) is satisfied only by C and G, and (v) is not satisfied by any of them.

Alternatively, another way to address the aggregation problem is to rely on a disaggregative measure of inequality such as a Lorenz curve. A problem here, though, is that two Lorenz curves being compared often, if not typically, cross, so a clear comparison is not straightforward.

The second basic criterion of conventional summary measures of inequality is the implicit value judgement problem. That is, any summary inequality measure involves implicit value judgements or weightings of different persons' incomes (or economic well-being), and thus contains embedded in it an implicit social welfare function (SWF). For example, different inequality measures differently emphasize income differences at the bottom, middle, or upper end of the distribution. Consequently, it can be argued, it would be better to choose desirable SWF properties explicitly and then derive the implied inequality measure from the desired SWF. To do so is to take a *normative* approach to measuring inequality rather than the traditional descriptive approach. This is the approach taken in the current paper.

To implement such a normative approach, one first needs to define a social welfare function and its basic properties. For a much more expansive discussion of the normative approach, see, for example, Boadway and Bruce (1984). Specifically, a social welfare function $W(\bullet)$ is any function

$$W = f(U_1, \ldots, U_N)$$

that has as arguments U_i individual (or household) utility functions and that incorporates social values used to aggregate economic well-being across the population. To do this, we require that:

- the U_i's must be at least cardinal scale measurable in order to be aggregated across persons;
- the U_i's must have at least some degree of comparability across persons in the population (i.e., if utilities are cardinally measured for each individual, the units of measurement must be the same across individuals); and
- for technical convenience, each U_i depends only on incomes and indeed only on individual *i*'s income (i.e., $U_i = U_i(Y_i)$, so there is no envy or altruism). This implies an additively separable SWF, and each $U_i(\bullet)$ is viewed as a "social income valuation function".

One can then identify several possible desirable properties for such a social welfare function:

i) (Strong) Pareto Principle – State *X* is socially preferred to state *Y* if at least one person strictly prefers *X* to *Y* and no one prefers state *Y* to *X* (i.e., $\partial U_i \,/\, \partial Y_i > 0$ and social indifference curves in Y_i, Y_j space have negative slopes);

ii) Symmetry or Anonymity – Everyone's incomes are evaluated by using the same $U(\bullet)$ function (i.e., $U_i(\bullet) = U(\bullet)$ for all $i = 1, \ldots, N$);

iii) Population Invariance – If the population is replicated *K* times, then social welfare increases *K*-fold (i.e., $W(Y_1, \ldots, Y_{KN}) = K \bullet W(Y_1, \ldots, Y_N))$;

iv) Strict Concavity of the SWF or the Principle of Transfers – A strictly concave SWF is such that $\partial^2 U_i \,/\, \partial Y_i^2 < 0$ for all *i* (this implies that social indifference curves are strictly convex to the origin). This is sometimes referred to as an "egalitarian SWF";

v) Transfer Sensitivity – A transfer-sensitive SWF is such that

$$\frac{\partial^3 U_i}{\partial Y_i^3} > 0 .$$

Again, this is a stronger version of the Principle of Transfers.

Atkinson (1971) uses this normative approach to show that, under properties (i)-(iv), an empirical proxy of social welfare or economic well-being (SW_p) can be expressed as

$$SW_p = \bar{Y} \bullet (1 - I_A)$$
$$= \bar{Y} \bullet \mathrm{E}$$

where $\bar{Y}$ is the mean income level of a distribution and I_A is a specific measure of inequality (referred to as Atkinson's inequality measure), where it turns out that $0 \leq I_A \leq 1$ where higher values indicate greater levels of inequality in the distribution. That is, SW_p can be decomposed into two (multiplicative) components – an efficiency dimension $(\bar{Y})$ or average per capita income and an equity dimension *(E)* where $\equiv 1 - I_A$.

If one further assumes a specific functional form for $U(\bullet)$ – in the convenient form of an iso-elastic social welfare function – Atkinson (1970) then derives a specific formula for the calculation of I_A . An iso-elastic SWF is general and flexible enough to incorporate a wide range of social attitudes to income inequality from the Benthanite utilitarian SWF to Rawls' maxi-min SWF.

But I_A is still a summary or aggregate measure of income inequality. What the social choice literature since Atkinson's (1970) paper has tried to do is to extend or apply Atkinson's normative perspective to develop a set of disaggregative criteria for comparing different income distributions based on the above properties, so that both criticisms of traditional inequality measures are addressed. The rest of this paper examines several such disaggregative criteria from the theoretical social choice literature and proposes ways to operationalize or empirically implement these criteria in terms of vectors of quantile means and income shares and related disaggregative distributional statistics. The paper also develops inference procedures to allow for formal statistical testing for these criteria. This development is applied to six such criteria in the following sections.

4. Application to Rank Dominance and a Practical Empirical Criterion

One early example of a disaggregative normative ranking criterion for distributions comes from Saposnik (1981). His rank dominance theorem says that, for any social welfare function satisfying the properties of symmetry, population invariance and the Pareto principle (i.e., social welfare conditions (i) – (iii)), distribution *A* is socially preferred to distribution *B* if the quantile means for *A* are all higher than those for *B*. Note that there is no egalitarianism built into this criterion. It essentially says that, if everyone has higher incomes in *A* than in *B*, then they must be better off. This is useful in comparing distributions many years apart, say, for example, the Canadian income distributions for 1961 versus 2021. But in most practical cases faced by empirical researchers, this situation doesn't apply.

Nonetheless, it is useful to begin our application of dominance criteria with this relatively simple criterion. To empirically implement it, one represents the two distributions being compared by their respective vectors of sample quantile means, $\hat{\mu}_i$, for $i = 1, \ldots, K$ quantiles. The actual decision rule for determining the outcome of the comparison of vectors requires some practical empirical criterion (henceforth a PEC) based on the principles of statistical inference.

4.1 A Practical Empirical Criterion for Quantile Means

Following Beach, Davidson and Slotsve (1984), one can set out a two-step test procedure for the PEC. It is assumed that the data samples for the two distributions being compared are independent and hence do not overlap. Examples are, say, two different years of data being compared or two different (non-overlapping) population groups such as age, racial, or sex groups.

Step 1 – Test the joint null hypothesis of equality of the two (population) quantile mean vectors versus the alternative hypothesis of non-equality. This can be done by a standard (but asymptotic) chi-square test with *K* degrees of freedom, where *K* is the number of quantiles. For a meaningful disaggregative analysis, it makes sense to let *K* = 10 or 20, say, rather than a small number such as 5. If the null hypothesis is not rejected, then the two distributions can be said to be not statistically significantly different, and further comparison is not pursued.

<u>Step 2</u> – If, however, the null hypothesis in Step 1 is rejected – which is the typical case when using large microdata sets for the sample distributions – then proceed to calculate separate t-statistics for differences for each of the individual quantile means. These *K* individual t-statistics, however, are correlated, and hence comparing each test statistic to the critical value on an (asymptotic) normal distribution would not be appropriate. One has to recognize that this Step 2 involves correlated multiple comparisons. Following the work of Beach and Richmond (1985) and Bishop, Formby and Thistle (1989, 1992) on multiple comparison testing, one should compare the *K* separate t-statistics (for differences in quantile means) to critical values on the Studentized Maximum Modules (or SMM) distribution. If at least one of the quantile mean differences t-statistics has the appropriate sign and is statistically significant (based on the SMM distribution) and none of the t-statistics of the remaining quantile mean differences has the wrong sign and is significant, then conclude that the distribution with the higher sample quantile means rank dominates (or is socially preferred to) that with the lower quantile means. If not, then one can say only the two distributions are statistically significantly different and not reach a preferred or dominance conclusion. Note that this is an asymptotic test and critical values from the SMM distribution correspond to *K* and infinite degrees of freedom. Typical usefully critical values from the SMM distribution are:

	$\propto = .01$	$\propto = .05$	$\propto = .10$
K = 5 –	3.289	2.800	2.560
K = 10 –	3.691	3.254	3.043
K = 20 –	4.043	3.643	3.453

Source: Stoline and Ury (1979), Tables 1-3.

4.2 Full Variance-Covariance Matrix for Quantile Means

The first step in the above practical empirical criterion (PEC) involves a joint test of the difference between two vectors or sets of estimated quantile means. If the two distributions being compared are designated distributions *A* and *B*, then the vectors of quantile means can be represented as

$$\hat{\mu}^a = (\hat{\mu}_1^a, \ldots, \hat{\mu}_K^a)' , \quad \mu^a = (\mu_1^a, \ldots, \mu_K^a)'$$

and $$\hat{\mu}^b = \left(\hat{\mu}_1^b, \ldots, \hat{\mu}_K^b\right)' , \quad \mu^b = \left(\mu_1^b, \ldots, \mu_K^b\right)' .$$

A standard result from statistics, then, shows that if the random vector $\hat{\mu}^a$ is normally distributed with mean μ^a and variance-covariance matrix V^a, $\hat{\mu}^b$ is normally distributed with mean μ^b and variance-covariance matrix V^b, and $\hat{\mu}^a$ and $\hat{\mu}^b$ are statistically independent, then $\hat{\mu}^b - \hat{\mu}^a$ is also normally distributed with mean $\mu^b - \mu^a$ and variance-covariance matrix $V^a + V^b$. Under the null hypothesis that the two vectors μ^a and μ^b are the same (i.e., $\mu^b - \mu^a = 0$), then the quadratic form

$$(\hat{\mu}^b - \hat{\mu}^a)'[V^a + V^b]^{-1}(\hat{\mu}^b - \hat{\mu}^a)$$

is distributed as a chi-squared random variable with K degrees of freedom. If V^a and V^b are estimated consistently, then the test statistics for step 1 of the PEC,

$$(\hat{\mu}^b - \hat{\mu}^a)'\left[\hat{V}^a + \hat{V}^b\right]^{-1}(\hat{\mu}^b - \hat{\mu}^a) \tag{18}$$

is asymptotically distributed as a chi-square variate again with *K* degrees of freedom.

In order to implement the chi-square test in (18), however, one needs to know how to estimate the full variance-covariance matrices $\hat{V}^a$ and $\hat{V}^b$ of $\hat{\mu}^a$ and $\hat{\mu}^b$, respectively. The development in Section 2 above showed how to obtain the estimated variances (the square of the estimated standard errors of the various individual quantile means). But, in order to perform the joint chi-square test in Step 1, one also needs estimates for all the covariances in $\hat{V}^a$ and $\hat{V}^b$ as well. The approach followed to obtain them, however, is the same as for the variances.

Argumentation is expressed in terms of asymptotic variances and covariances. Again, let $\hat{m} = (\hat{\mu}_1, \dots, \hat{\mu}_K)'$ generically represent the vector of sample quantile mean estimates for a given income distribution, so it has been shown that

$$Asy.var(\hat{m}) \equiv V_S = G\,\Lambda\,G' \tag{19}$$

where Λ is the asymptotic variance-covariance matrix of the sample quantile cut-off levels, the $\hat{\xi}_i$'s, and G is a *K x (K-1)* matrix of partial derivatives

$$G = [g_{ij}] \qquad \text{where } g_{ij} = \frac{\partial \mu_i}{\partial \xi_j} \tag{20a}$$

for $i = 1, \dots, K$ rows and $j = 1, \dots, K-1$ columns.

Note that, in this development, the asymptotic variance-covariance matrix of $\hat{m}$ is V_S with a subscript S (for asymptotic) to distinguish it from matrix $\hat{V}$ which refers to the estimated variance-covariance matrix of $\hat{m}$ which is gotten by rescaling the estimated asymptotic variance-covariance matrix by the inverse of the sample size. We also use a slightly more general notation than before with

$$\mu_i = \left(\frac{1}{D_i}\right) \bullet \int_{\xi_{i-1}}^{\xi_i} y\, f(y) dy$$

$$= \left(\frac{1}{D_i}\right) \bullet N_i(\xi_{i-1}, \xi_i) \qquad \text{for } i = 1, \dots, K$$

where $D_i = p_i - p_{i-1}$ and $\xi_0 = 0$.[3] Then

$$g_{ij} = \left(\frac{1}{D_i}\right) \bullet \frac{\partial N_i}{\partial \xi_j}. \tag{20b}$$

[3] This allows for different sized quantile groups.

The *i,j*'th element of V_S, then, is obtained by premultiplying the matrix Λ by the *i*'th row of *G* treated as a row vector and postmultiplying by the *j*'th row of *G* (written as a column vector):

$$v_s(i,j) = (i'\text{th row of } G) \bullet \Lambda \bullet (j'\text{th row of } G)' \,. \qquad (21)$$

In the case of variances $i = j$, so calculations lead to

$$Asy.var(\hat{\mu}_1) \equiv Asy.var(\hat{m}_1)$$

$$= p_1(1 - p_1)\left(\frac{\xi_1}{D_1}\right)^2 \qquad (22a)$$

and $$Asy.var(\hat{\mu}_K) \equiv Asy.var(\hat{m}_K)$$

$$= p_{K-1}(1 - p_{K-1})\left(\frac{\xi_{K-1}}{D_K}\right)^2 \;. \qquad (22b)$$

For *i = 2, ..., K-1*:

$$Asy.var(\hat{\mu}_i) = p_{i-1}(1 - p_{i-1})\left(\frac{\xi_{i-1}}{D_i}\right)^2 + p_i(1 - p_i)\left(\frac{\xi_i}{D_i}\right)^2$$

$$- 2p_{i-1}(1 - p_i)\left(\frac{\xi_{i-1}\xi_i}{{D_i}^2}\right) \;. \qquad (22c)$$

Equations (22a)-(22c) determine the elements on the principal diagonal of V_S .

Now consider the off-diagonal elements in the first row of V_S . For $1 < j < K$:

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_j) = -p_1(1 - p_{j-1})\left(\frac{\xi_1}{D_1}\right)\left(\frac{\xi_{j-1}}{D_j}\right) + p_1(1 - p_j)\left(\frac{\xi_1}{D_1}\right)\left(\frac{\xi_j}{D_j}\right), \qquad (22d)$$

For elements along the last column of V_S , i.e., for $1 < i < K$:

$$Asy.cov(\hat{\mu}_i, \hat{\mu}_K) = p_{i-1}(1 - p_{K-1})\left(\frac{\xi_{i-1}}{D_i}\right)\left(\frac{\xi_{K-1}}{D_K}\right) + p_i(1 - p_{K-1})\left(\frac{\xi_i}{D_i}\right)\left(\frac{\xi_{K-1}}{D_K}\right). \qquad (22e)$$

For the top right-hand corner element,

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_K) = -p_1(1 - p_{K-1})\left(\frac{\xi_1}{D_1}\right)\left(\frac{\xi_{K-1}}{D_K}\right). \qquad (22f)$$

For all remaining above-diagonal elements of V_S ; i.e., for $1 < i < j < K$:

$$Asy.cov(\hat{\mu}_i, \hat{\mu}_j) = p_{i-1}(1 - p_{j-1})\left(\frac{\xi_{i-1}}{D_i}\right)\left(\frac{\xi_{j-1}}{D_j}\right) - p_{i-1}(1 - p_j)\left(\frac{\xi_{j-1}}{D_i}\right)\left(\frac{\xi_j}{D_j}\right)$$

$$- p_i(1 - p_{j-1}) \left(\frac{\xi_i}{D_i}\right)\left(\frac{\xi_{j-1}}{D_j}\right) + p_i(1 - p_j) \left(\frac{\xi_i}{D_i}\right)\left(\frac{\xi_j}{D_j}\right). \tag{22g}$$

Since a variance-covariance matrix is symmetric about its principal diagonal, all below-diagonal elements can be obtained as

$$v_S(i,\ j) = v_S(j,\ i) \qquad \text{for } i > j . \tag{22h}$$

Note also that all terms in the V_S matrix – both (asymptotic) variances and covariances – are also distribution-free in that they do not depend on $f(\bullet)$ evaluations, and thus can all be readily estimated consistently. Thus consistent estimates of the actual variances and covariances of the μ_i's can be obtained as

$$\hat{v}(i,\ j) = \hat{v}_S(i,\ j)\ /\ N \tag{23}$$

where *N* is the size of the estimation sample.

In order to perform Step 1 of the PEC for comparing the two quantile mean vectors $\hat{\mu}^a$ and $\hat{\mu}^b$, then, calculate estimates of all asymptotic variances and covariances ($\hat{V}_S^a$ and $\hat{V}_S^b$) for the two samples using the formulas in equations (22a)-(22h) by replacing population parameters by their consistent sample estimates, rescale the (asymptotic) variance-covariance estimates to the actual variance-covariance estimates $\hat{V}^a$ and $\hat{V}^b$ as in (23), and then calculate the joint chi-square test statistic in (18).

To perform the individual tests in Step 2 of the PEC, compute the standard "t-statistic" ratio for the difference between two independent random variables ($\hat{\mu}_i^a$ and $\hat{\mu}_i^b$) as

$$t_i = \frac{\hat{\mu}_i^b - \hat{\mu}_i^a}{\left[\hat{v}^a(i,i) + \hat{v}^b(i,i)\right]^{1/2}}$$

and compare this to the appropriate critical value on the SMM distribution.

5. Application to Lorenz Dominance

The same approach can be applied to an inequality-based dominance criterion. Atkinson, in his famous 1970 paper, forwarded what has come to be known as the Lorenz dominance theorem. For any (summary) inequality measure satisfying symmetry, mean independence, population homogeneity and the principle of transfers (i.e., inequality criteria (i)-(iv) above), if the Lorenz curve for distribution A lies everywhere above the Lorenz curve for distribution B, then all inequality measures satisfying these properties will indicate that (summary) inequality in A is less than in B. Note that this theorem does not say anything about social welfare; it refers only to inequality. It also does not say anything if the two Lorenz curves cross.

To empirically implement this dominance criterion, one can represent a Lorenz curve by a vector of its estimated ordinates. Testing between Lorenz curves then amounts to tests of differences between the estimated ordinate vectors. Again, if the two distributions whose inequality is being compared are designated A and B, then the vectors of Lorenz curve ordinates can be represented by

$$\hat{l}^a = \left(\hat{l}_1^a, \dots, \hat{l}_{K-1}^a\right)', \quad l^a = (l_1^a, \dots, l_{K-1}^a)'$$

and $$\hat{l}^b = \left(\hat{l}_1^b, \dots, \hat{l}_{K-1}^b\right)', \quad l^b = \left(l_1^b, \dots, l_{K-1}^b\right)',$$

and their respective variance-covariance matrices by Φ^a and Φ^b . The ordinates $l_1, \dots, l_{K-1}$ correspond to the given (cumulative) proportions $p_1, \dots, p_{K-1}$. Since the two end points on a Lorenz curve are fixed at $p_0 = 0$ and $p_K = 1$, only *K-1* ordinates are random variables.

The actual decision rule or PEC for comparing the vectors of Lorenz curve ordinates again involves two steps. And again it is assumed that the two sets of ordinate estimates are statistically independent and based on two quite separate samples.

Step 1 – Test the joint null hypothesis of equality of the two ordinate vectors (i.e., $l^b - l^a = 0$) versus the alternative hypothesis of non-equality. In this case, the test statistic is

$$(\hat{l}^b - \hat{l}^a)'\left[\hat{\Phi}^a + \hat{\Phi}^b\right]^{-1}(\hat{l}^b - \hat{l}^a) \qquad (24)$$

which is distributed asymptotically as a chi-square random variable with *K-1* degrees of freedom. If the null hypothesis is not rejected, then the two Lorenz curves can be said to be not statistically significantly different, and further comparison is not pursued.

Step 2 – If, however, the null hypothesis in Step 1 is rejected, then undertake separate "t-statistic" calculations for differences on each of the individual estimated Lorenz curve ordinates. If at least one of the t-statistics has the appropriate sign and is statistically significant compared to critical values on the SMM distribution with *K-1* and infinite degrees of freedom and none of the t-statistics (if any) that has the wrong sign is statistically significant (again based on the SMM distribution), then one can conclude that one set of ordinates statistically dominates the other. If statistical dominance is found, this implies dominance for all summary inequality measures satisfying inequality properties (i)-(iv). Again, typical useful SMM critical values are:

	$\propto = .01$	$\propto = .05$	$\propto = .10$
K-1 = 4	3.430	2.631	2.378
K-1 = 9	3.634	3.190	2.976
K-1 = 19	4.018	3.615	3.425

Source: Stoline and Ury (1979), Tables 1-3.

This leaves two problems to be resolved: (i) how does one determine the statistical properties of the Lorenz curve ordinates in order to make statistical inference decisions, and (ii) how to establish the full variance-covariance matrix of the vector of estimated ordinates. These are addressed in the next two subsections.

5.1 Inference for Lorenz Curve Ordinates

Recall that Lorenz curve ordinates are simply cumulative income shares (which have already been considered in Section 2 above). Let the K-vector of individual income share statistics be

$$\hat{n} = (\hat{n}\cdot, \ldots, \hat{n}_K)'$$

with corresponding population shares $n = (n_1, \ldots, n_K)'$. Then it can be seen that

$$\hat{l} = U \bullet \hat{n} \qquad (25)$$

where U is a $(K-1)\ x\ K$ matrix with ones on its principal diagonal and below, and zeros above the diagonal. U is given and non-random. Since (25) is a linear transformation, if $\hat{n}$ is (asymptotically) joint normally distributed with mean vector n and variance-covariance matrix W_S, then $\hat{l}$ is also (asymptotically) joint normally distributed with mean $l = U \bullet n$ and asymptotic variance-covariance matrix

$$\Phi_S = U \bullet W_S \bullet U' \qquad \text{and hence} \qquad \Phi = U \bullet W \bullet U' . \qquad (26)$$

So if *W*, the actual variance-covariance matrix of the estimates income shares can be established, so also can that of the vector of implied Lorenz curve ordinates.

5.2 Full Variance-Covariance Matrix for Income Shares

In order to obtain estimates of variance-covariance matrix elements for sample income shares, it is again useful to work out asymptotic variances and covariances for $\hat{n}$. It has been established in Section 2 that

$$Asy.var(\hat{n}_i) = G_i' \, \Sigma_i \, G_i$$

where Σ_i is the asymptotic variance-covariance matrix of the triplet $\hat{\xi}_{i-1}$, $\hat{\xi}_i$, and $\hat{\mu}$ and

$$G_i = \left[\frac{\partial N_i}{\partial \xi_{i-1}}, \frac{\partial N_i}{\partial \xi_i}, \frac{\partial N_i}{\partial \mu}\right]'$$

where income share $n_i \equiv IS_i = \int_{R_i} \left(\frac{1}{\mu}\right) y\, f(y) dy \equiv N_i\,(\hat{\xi}_{i-1}, \hat{\xi}_i, \mu)$. More generally, one can use a multivariate version of the Rao linkage theorem to establish that

$$Asy.var(\hat{n}) \equiv W_S = [w_S(i,j)] = G'\,\Sigma\, G \tag{27}$$

where now Σ is the *KxK* asymptotic variance-covariance matrix of the full set of sample quantile cut-offs, the $\hat{\xi}_i$'s, and the overall sample mean, $\hat{\mu}$. Σ thus consists of **Λ** in the upper-left *K-1* rows and columns, the (asymptotic) covariances of the $\hat{\xi}_i$ and $\hat{\mu}$ along the bottom row and right-hand column, and the (asymptotic) variance of $\hat{\mu}$ $(= \sigma^2)$ in the bottom right-hand corner. The *KxK* matrix *G* of partial derivatives then has as its *i*'th row all zeros except for the three elements $\frac{\partial N_i}{\partial \xi_{i-1}}, \frac{\partial N_i}{\partial \xi_i}, \frac{\partial N_i}{\partial \mu}$. Thus it is more convenient to work out the terms of W_S element by element where $w_S(i,j) = (i\text{'th row of } G) \bullet \Sigma \bullet (j'\text{th row of } G)'$. (28)

In the case of variances, *i=j*, which works out to the results:

$$Asy.var(\widehat{IS}_1) \equiv Asy.var(\hat{n}_1) = w_S(1,1)$$

$$= p_1(1-p_1)\left(\frac{\xi_1}{\mu}\right)^2 + \left(\frac{IS_1}{\mu}\right)^2 \sigma^2 \tag{29a}$$

$$-2\left(\frac{\xi_1}{\mu}\right)\left(\frac{IS_1}{\mu}\right)[\xi_1 - \mu(1-p_1)]\,,$$

$$Asy.var(\widehat{IS}_K) \equiv Asy.var(\hat{n}_K) = w_S(K,K)$$

$$= p_{K-1}(1-p_{K-1})\left(\frac{\xi_{K-1}}{\mu}\right)^2 + \left(\frac{IS_K}{\mu}\right)^2 \sigma^2 \tag{29b}$$

$$+2\left(\frac{\xi_{K-1}}{\mu}\right)\left(\frac{IS_K}{\mu}\right)[\xi_{K-1} - \mu(1-p_{K-1})]\,.$$

And for *i = 2, ..., K-1*:

$$Asy.var(\widehat{IS}_i) \equiv Asy.var(\hat{n}_i) = w_S(i,i)$$

$$= p_{i-1}(1-p_{i-1})\left(\frac{\xi_{i-1}}{\mu}\right)^2 + p_i(1-p_i)\left(\frac{IS_i}{\mu}\right)^2 + \left(\frac{IS_i}{\mu}\right)^2 \sigma^2$$

$$-2\ p_{i-1}(1-\ p_i)\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{\xi_i}{\mu}\right) \tag{29c}$$

$$+2\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{IS_i}{\mu}\right)\ [\xi_{i-1}-\ \mu(1-\ p_{i-1})]$$

$$-2\left(\frac{\xi_i}{\mu}\right)\left(\frac{IS_i}{\mu}\right)\ [\xi_i-\ \mu(1-\ p_i)]\ .$$

Now address the (asymptotic) covariances in the first row of W_S . For $1\ <j\ <K$:

$$Asy.cov\left(\widehat{IS}_1,\widehat{IS}_j\right)=\ w_S(1,j)$$

$$=\ -p_1\left(1-\ p_{j-1}\right)\left(\frac{\xi_1}{\mu}\right)\left(\frac{\xi_{j-1}}{\mu}\right)+\ p_1\left(1-\ p_j\right)\left(\frac{\xi_1}{\mu}\right)\left(\frac{\xi_j}{\mu}\right)+\left(\frac{IS_1}{\mu}\right)\left(\frac{IS_j}{\mu}\right)\ \sigma^2$$

$$-\ \left(\frac{\xi_1}{\mu}\right)\left(\frac{IS_j}{\mu}\right)\ [\xi_1-\ \mu(1-\ p_1)] \tag{29d}$$

$$+\left(\frac{\xi_{j-1}}{\mu}\right)\left(\frac{IS_1}{\mu}\right)\ \left[\xi_{j-1}-\ \mu(1-\ p_{j-1})\right]$$

$$-\left(\frac{\xi_j}{\mu}\right)\left(\frac{IS_1}{\mu}\right)\ \left[\xi_j-\ \mu(1-\ p_j)\right]\ .$$

For elements down the last column of W_S ; i.e., for $1\ <i\ <K$:

$$Asy.cov\left(\widehat{IS}_i,\widehat{IS}_K\right)=\ w_S(i,K)$$

$$=\ p_{i-1}(1-\ p_{K-1})\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{\xi_{K-1}}{\mu}\right)-\ p_i(1-\ p_{K-1})\left(\frac{\xi_i}{\mu}\right)\left(\frac{\xi_{K-1}}{\mu}\right)+\left(\frac{IS_i}{\mu}\right)\left(\frac{IS_K}{\mu}\right)\ \sigma^2$$

$$+\ \left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{IS_K}{\mu}\right)\ [\xi_{i-1}-\ \mu(1-\ p_{i-1})] \tag{29e}$$

$$-\left(\frac{\xi_i}{\mu}\right)\left(\frac{IS_K}{\mu}\right)\ [\xi_i-\ \mu(1-\ p_i)]\ .$$

$$+\left(\frac{\xi_{K-1}}{\mu}\right)\left(\frac{IS_i}{\mu}\right)\ [\xi_{K-1}-\ \mu(1-\ p_{K-1})]\ .$$

For the top right-hand corner element of W_S,

$$Asy.cov\left(\widehat{IS}_1,\widehat{IS}_K\right)=\ w_S(1,K)$$

$$=\ -p_1(1-\ p_{K-1})\left(\frac{\xi_1}{\mu}\right)\left(\frac{\xi_{K-1}}{\mu}\right)+\ \left(\frac{IS_1}{\mu}\right)\left(\frac{IS_K}{\mu}\right)\ \sigma^2$$

$$-\ \left(\frac{\xi_1}{\mu}\right)\left(\frac{IS_K}{\mu}\right)\ [\xi_1-\ \mu(1-\ p_1)] \tag{29f}$$

$$+\left(\frac{\xi_{K-1}}{\mu}\right)\left(\frac{IS_1}{\mu}\right)\,[\xi_{K-1}-\mu(1-p_{K-1})]\,.$$

Finally, for all remaining above-diagonal elements of W_S ; i.e., for $1 < i < j < K$:

$$Asy.cov(\widehat{IS}_i, \widehat{IS}_j) = w_S(i,j)$$

$$= p_{i-1}(1-p_{j-1})\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{\xi_{j-1}}{\mu}\right) - p_{i-1}(1-p_j)\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{\xi_j}{\mu}\right) + \left(\frac{IS_i}{\mu}\right)\left(\frac{IS_j}{\mu}\right)\sigma^2$$

$$- p_i(1-p_{j-1})\left(\frac{\xi_i}{\mu}\right)\left(\frac{\xi_{j-1}}{\mu}\right) + p_i(1-p_j)\left(\frac{\xi_i}{\mu}\right)\left(\frac{\xi_j}{\mu}\right)$$

$$+\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{IS_j}{\mu}\right)\,[\xi_{i-1}-\mu(1-p_{i-1})] \qquad (29g)$$

$$-\left(\frac{\xi_i}{\mu}\right)\left(\frac{IS_j}{\mu}\right)\,[\xi_i-\mu(1-p_i)]$$

$$+\left(\frac{\xi_{j-1}}{\mu}\right)\left(\frac{IS_i}{\mu}\right)\,[\xi_{j-1}-\mu(1-p_{j-1})]$$

$$-\left(\frac{\xi_j}{\mu}\right)\left(\frac{IS_i}{\mu}\right)\,[\xi_j-\mu(1-p_j)]\,.$$

Again, since a variance-covariance matrix is symmetric about its principal diagonal, all below-diagonal covariance terms can be obtained as

$$w_S(i,j) = w_S(j,i) \qquad \text{for } i > j\,. \qquad (29h)$$

Note also that all terms in W_S are distribution-free, and thus can be readily estimated consistently. More specifically, consistent estimates of the actual variances and covariances of the $\widehat{IS}_i$ can thus be obtained as

$$\widehat{w}(i,j) = \widehat{w}_S(i,j)\,/\,N \qquad (30)$$

where *N* is the estimation sample size.

Once again to perform Step 1 of the PEC for comparing the two vectors of Lorenz curve ordinates $\hat{l}^a$ and $\hat{l}^b$, first calculate estimates of all the asymptotic variances and covariances ($\widehat{W}_S^a$ and $\widehat{W}_S^b$) for the two estimation samples from equations (29a)-(29h) by replacing population parameters by their consistent sample estimates, rescale the (asymptotic) variance and covariance estimates to the actual variance and covariance estimates ($\widehat{W}^a$ and $\widehat{W}^b$) as in equation (30), calculate the Lorenz curve ordinates by $\hat{l} = U \bullet \hat{n}$ from equation (25) and Lorenz curve ordinate estimated variances and covariances from

$$\widehat{\Phi} = U \bullet \widehat{W} \bullet U'$$

following equation (26), and then finally calculate the joint chi-square test statistic in equation (24).

To perform the individual tests in Step 2 of the PEC, again use the standard "t-statistic" ratio for the difference between two independent random variates ($\hat{l}_i^a$ and $\hat{l}_i^b$) as

$$t_i = \frac{\hat{l}_i^b - \hat{l}_i^a}{\left[\widehat{w}^a(i,i) + \widehat{w}^b(i,i)\right]^{1/2}}$$

and compare this to the relevant critical value on the SMM tables.

6. Application to Generalized Lorenz Dominance

A blending of the first two dominance criteria is provided in a third application of empirically implementing curve-based dominance criteria. Shorrocks (1983) uses a transformed Lorenz curve as the basis for social welfare inferences, not just inequality conclusions. The generalized Lorenz dominance theorem of Shorrocks (1983) says that, for any additively separable social welfare function satisfying social welfare conditions (i)-(iv) including the principle of transfers, distribution *A* is socially preferred to distribution *B* if the generalized Lorenz curve for *A* lies everywhere above the generalized Lorenz curve for *B*. The generalized Lorenz curve ordinates for an income distribution are obtained by scaling up the Lorenz curve ordinates of the distribution by the distribution's overall mean income level:

$$g_i = \mu \bullet l_i \quad \text{and } \hat{g}_i = \hat{\mu} \bullet \hat{l}_i \,. \tag{31}$$

Essentially, the argument is that, if the mean income of the distribution *A* is sufficiently higher than that in distribution *B*, this can compensate for some greater degree of inequality in *A* than in *B*, so that social welfare will still be greater in distribution *A* than in *B*. It turns out that this rule is very convenient for ranking social welfare among quite disparate countries, or for ranking income distributions in a given country (or group) over long periods of time (e.g., the Canadian income distribution across the decades of 1950, 1960, 1970 and 1980).

To implement this dominance criterion, one can again represent a generalized Lorenz curve by a vector of its estimated ordinates:

$$g = (g_1, \dots, g_{K-1})' \qquad \text{and } \hat{g} = (\hat{g}_1, \dots, \hat{g}_{K-1})' \,.$$

Testing between generalized Lorenz curves then amounts to tests of differences between the estimated ordinate vectors $\hat{g}^a$ and $\hat{g}^b$. The respective generalized Lorenz curve ordinate variance-covariance matrices may be labelled Ψ^a and Ψ^b .

The corresponding decision rule or PEC for comparing vectors $\hat{g}^a$ and $\hat{g}^b$ once again involves two steps (where, as above) the estimation samples are statistically independent.

Step 1 – Test the joint null hypothesis of equality of two generalized Lorenz curve ordinate vectors (i.e., $g^b - g^a = 0$) versus the alternative hypothesis of non-equality. In this case, the test statistics is

$$(\hat{g}^b - \hat{g}^a)' \left[\hat{\Psi}^a + \hat{\Psi}^b\right]^{-1} (\hat{g}^b - \hat{g}^a) \tag{32}$$

which, under the null hypothesis, is asymptotically distributed as a chi-square random variable with *K-1* degrees of freedom. If the null hypothesis is not rejected, then the two generalized Lorenz curves can be said to be not statistically significantly different, and further comparison is not warranted.

Step 2 – If, however, the null hypothesis in Step 1 is rejected, then proceed to compute separate "t-statistics" for differences on each of the individual generalized Lorenz curve ordinates. If at least one of the t-statistics has the appropriate sign and is statistically significant compared to critical values on the SMM distribution with *K-1* and infinite degrees of freedom and none of the t-statistics (if any) that have the wrong sign is statistically significant (again on the SMM critical values), then one can conclude that the distribution with the higher sample generalized Lorenz curve ordinates rank dominates (or is socially preferred to) the distribution with the corresponding lower ordinates. If not, then one can say only that the social welfare of the two distributions are statistically significantly different, but not reach a preferred or dominance conclusion.

6.1 Variance-Covariance Structure of Generalize Lorenz Curve Ordinates

Since Lorenz curve ordinates are calculated from income shares, it makes sense to consider the relationship of generalized Lorenz curve ordinates to these underlying income shares as well. To go back to first principles, consider $\mu \bullet IS_i$ as the dollar contribution of the i'th income group to the overall mean income of the distribution. So we can represent it by the "contribution"

$$c_i = \mu \bullet IS_i \tag{33a}$$

$$= \mu \bullet \int_{R_i} \left(\frac{1}{\mu}\right) y\, f(y)dy$$

$$= \int_{R_i} y\, f(y)dy \qquad \equiv \quad N_i\,(\xi_{i-1},\ \xi_i)$$

$$= D_i \bullet \left(\frac{1}{D_i}\right) N_i(\xi_{i-1},\ \xi_i)$$

$$= D_i \bullet \mu_i \tag{33b}$$

where μ_i is the quantile mean of the i'th income group and $D_i = p_i - p_{i-1}$. So c_i is simply a scalar transform of the quantile mean μ_i . Consequently, $\hat{c}_i = D_i \bullet \hat{\mu}_i$, and the elements of the variance-covariance matrix of the vector of sample contributions $\hat{c} = (\hat{c}_1, \ldots, \hat{c}_K)'$ are simply scalar transforms of the corresponding elements of the variance-covariance matrix of the vector of quantile means $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_K)'$. More specifically, all the D_i terms in equations (22a)-(22h) for the asymptotic variances and covariances drop out. Alternatively viewed, since $c_i = N_i(\xi_{i-1}, \xi_i)$, all the partial derivatives in equations (20a) and (20b) now involve simply $\frac{\partial N_i}{\partial \xi_j}$ without the term $\frac{1}{D_i}$. For ease of reference, these may be set out explicitly for the asymptotic variance-covariance matrix Γ_S of $\hat{c}$:

$$Asy.var(\hat{c}_1) = p_1(1 - p_1)\xi_1^2 \tag{34a}$$

$$Asy.var(\hat{c}_K) = p_{K-1}(1 - p_{K-1})\xi_{K-1}^2 \tag{34b}$$

$$Asy.var(\hat{c}_i) = p_{i-1}(1 - p_{i-1})\xi_{i-1}^2 + p_i(1 - p_i)\xi_i^2 - 2\,p_{i-1}(1 - p_i)\xi_{i-1}\xi_i \qquad \text{for } i = 2, \ldots, K\text{-}1 . \tag{34c}$$

For $1 < j < K$:

$$Asy.cov(\hat{c}_1, \hat{c}_j) = -\,p_1(1 - p_{j-1})\xi_1\xi_{j-1} + p_1(1 - p_j)\xi_1\xi_j . \tag{34d}$$

For $1 < i < K$:

$$Asy.cov(\hat{c}_i, \hat{c}_K) = p_{i-1}(1 - p_{K-1})\xi_{i-1}\xi_{K-1} - p_i(1 - p_{K-1})\xi_i\xi_{K-1} . \tag{34e}$$

$$Asy.cov(\hat{c}_1, \hat{c}_K) = -\,p_1(1 - p_{K-1})\xi_1\xi_{K-1} . \tag{34f}$$

For $1 < i < j < K$:

$$Asy.cov(\hat{c}_i, \hat{c}_j) = p_{i-1}(1 - p_{j-1})\xi_{i-1}\xi_{j-1} - p_{i-1}(1 - p_j)\xi_{i-1}\xi_j - p_1(1 - p_{j-1})\xi_i\xi_{j-1} + p_i(1 - p_j)\xi_i\xi_j . \tag{34g}$$

And for all below-diagonal elements of $\Gamma_S = [\gamma_S(i,j)]$,

$$\gamma_S(i,j) = \gamma_S(j,i) \qquad \text{for } i > j\,. \tag{34h}$$

Thus consistent estimates of the actual variances and covariances of the $\hat{c}_i$'s can be obtained as

$$\hat{\gamma}(i,j) = \hat{\gamma}_S(i,j)\,/\,N \tag{35}$$

where here *N* is the estimation sample size.

The ordinates of the generalized Lorenz curve can be readily obtained from the $\hat{c}_i$ by straightforward cumulation:

$$\hat{g}_i = \textstyle\sum_{j=1}^{i} \hat{c}_j \qquad \text{and} \qquad g_i = \textstyle\sum_{j=1}^{i} c_j$$

or more generally,

$$\hat{g} = U \bullet \hat{c} \qquad \text{and} \qquad g = U \bullet c \tag{36}$$

where again *U* is a *(K-1)xK* non-random matrix with ones on the principal diagonal and below, and zeros above the diagonal.

Since the $\hat{c}_i$'s are proportional functions of the $\hat{\mu}_i$'s , and the $\hat{\mu}_i$'s are asymptotically joint normal, then $\hat{c}_i$'s are also asymptotically joint normal with means c_i's and full (asymptotic) variance-covariance matrix Γ_S given by equations (34a)-(34h) and estimated actual variance-covariance matrix $\hat{\Gamma}$. Similarly, since the $\hat{g}_i$'s are linear functions of the $\hat{c}_i$'s , the $\hat{g}_i$'s are also asymptotically joint normally distributed with means g_i's and full (asymptotic) variance-covariance matrix

$$\Psi_S = U \bullet \Gamma_S \bullet U' \tag{37}$$

and estimated actual variance-covariance matrix

$$\hat{\Psi} = U \bullet \hat{\Gamma} \bullet U' \tag{38}$$

where the elements in $\hat{\Gamma}$ are given by equation (35). Once again, all terms in Γ_S are distribution-free, and thus can be readily estimated consistently.

To perform Step 1 of the PEC for comparing the two vectors of generalized Lorenz curve ordinates $\hat{g}^a$ and $\hat{g}^b$, first calculate estimates of all the asymptotic variances and covariances ($\hat{\Gamma}_S^a$ and $\hat{\Gamma}_S^b$) for the two estimation samples from equations (34a)-(34h) by replacing population parameters by their consistent sample estimates, rescale the (asymptotic) variance and covariance estimates to actual variance and covariance estimates ($\hat{\Gamma}^a$ and $\hat{\Gamma}^b$) as in equation (35), calculate the generalized Lorenz curve ordinates by $\hat{g} = U \bullet \hat{c}$ from equation (36) and the generalized Lorenz curve ordinate estimated variances and covariances from equation (38), and then finally calculate the joint chi-square test statistic in equation (32).

To perform the individual tests in Step 2 of the PEC, again use the standard "t-statistic" ratio for the difference between two independent variates ($\hat{g}_i^a$ and $\hat{g}_i^b$) as

$$t_i = \frac{\hat{g}_i^b - \hat{g}_i^a}{\left[\hat{\gamma}^a(i,i) + \hat{\gamma}^b(i,i)\right]^{1/2}}$$

and compare this to the relevant critical value on the SMM distribution.

7. Decomposition of Generalized Lorenz Curve Ordinates into Efficiency and Equity Components

As stated in Section 3 above, Atkinson's inequality index, I_A , has the property – under various conditions – that an empirical proxy for social welfare can be decomposed into the product of an efficiency measure and an equity indicator:

$$\widehat{SW} = \hat{\mu} \bullet (1 - \hat{I}_A) .$$

A similar decomposition appears in Jorgenson (1990) as well. Can such an intuitively appealing decomposition also be applied more generally to entire dominance condition curves? The answer is yes. If the $\hat{\mu}_i$'s can be viewed as a disaggregative indicator of economic well-being or social welfare, then

$$\hat{\mu}_i = \hat{\mu} \bullet \left(\frac{\hat{\mu}_i}{\hat{\mu}}\right) \qquad \text{or alternatively} \quad = \hat{\mu}\left[1 - \left(\frac{\hat{\mu} - \hat{\mu}_i}{\hat{\mu}}\right)\right] \tag{39}$$

where $\hat{\mu}$ is again a measure of overall efficiency and $\left(\frac{\hat{\mu}_i}{\hat{\mu}}\right)$ can be viewed as an indicator of disaggregative equity for income group *i*. The term $\left(\frac{\mu_i}{\mu}\right)$ may be referred to as the relative-mean income gap for quantile group *i* and a vector of such terms as the relative-mean income curve for an income distribution (see Beach, 2021a, for further discussion and interpretation of this curve).

When comparing two income distributions, say A and B, it is obviously of interest to look at their differences in overall means, $\hat{\mu}^a$ and $\hat{\mu}^b$. But it is also of interest to consider the relative-mean income gaps across the various quantile groups and how these differ between the distributions. That is, consider the differences in the gaps, say, $\frac{\hat{\mu}_i^b}{\hat{\mu}^b} - \frac{\hat{\mu}_i^a}{\hat{\mu}^a}$, across all of the individual quantile groups as a reflection of the disaggregative equity differences between the two distributions.

Indeed, it turns out that performing formal statistical tests of these relative-mean income gaps is quite straightforward using the above development. For convenience, designate the relative-mean income gap for quantile group *i* by r_i . Then, from first principles,

$$\begin{aligned} r_i &= \left(\frac{\mu_i}{\mu}\right) \\ &= D_i^{-1} \bullet \int_{R_i} y\, f(y) dy \,/\, \mu \\ &= D_i^{-1} \bullet \int_{R_i} \left(\frac{y}{\mu}\right) f(y) dy \\ &= \left(\frac{1}{D_i}\right) \bullet IS_i = \left(\frac{1}{D_i}\right) \bullet n_i \end{aligned} \tag{40}$$

where, as before, $D_i = p_i - p_i$ and R_i is the appropriate range of integration. That is, r_i is simply a scalar transform of quantile *i*'s income share. And similarly,

$$\hat{r}_i = \left(\frac{\hat{\mu}_i}{\hat{\mu}}\right) = \left(\frac{1}{D_i}\right) \bullet \widehat{IS}_i = \left(\frac{1}{D_i}\right) \bullet \hat{n}_i \ .$$

Thus the vector $r = (r_1, \ldots, r_K)$ is such that

$$r = D^{-1} \bullet n \qquad \text{and similarly } \hat{r} = D^{-1} \bullet \hat{n} \tag{41}$$

where D^{-1} is a *KxK* matrix with elements D_i^{-1} along its principal diagonal and zeros elsewhere. Thus, since $\hat{r}$ is a linear transform of $\hat{n}$ and $\hat{n}$ is asymptotically joint normally distributed, so also is $\hat{r}$ with mean r and asymptotic variance-covariance matrix

$$R_S = D^{-1} \bullet W_S \bullet D^{-1} \tag{42}$$

where W_S is the asymptotic variance-covariance matrix of the estimated income share vector $\hat{n}$. Thus the asymptotic variance of $\hat{r}_i$ (for *i=2, …, K-1*), for example, is given by equation (29c) where each term is divided by D_i^2 . A consistent estimate of the actual variance-covariance matrix of $\hat{r}$ is then given by

$$\hat{R} = D^{-1} \bullet \widehat{W} \bullet D^{-1} = [\hat{r}(i,j)] \tag{43}$$

and the elements of $\widehat{W}$ are given by equation (30).

An asymptotic test of the difference in relative-mean income gaps for quantile group *i* between two independent distributions *A* and *B*, then, is done with the standard "t-statistic"

$$t_i = \frac{\hat{r}_i^b - \hat{r}_i^a}{\left[\hat{r}^a(i,i) + \hat{r}^b(i,i)\right]^{1/2}}$$

and this statistic is then compared to a critical value on the standard normal distribution. Note that the SMM distribution critical values are not used here since this test is not part of a PEC joint test criterion.

8. Inequality Dominance with a Single Lorenz Curve Crossing

What can one infer if Lorenz curves cross? More often than not this is the empirical situation when comparing two estimated Lorenz curves, and the Lorenz dominance criterion above is of no help in such situations. However, Shorrocks and Foster (1987) have come up with an extension of the latter criterion to cover just such situations. What may be called the transfer sensitivity dominance theorem states that, if the Lorenz curve for distribution *A* crosses the Lorenz curve for distribution *B* once from above, then all inequality measures satisfying the inequality properties (i)-(iv) plus property (v) – transfer sensitivity – will indicate that (summary) inequality in *A* is less than in *B* if the coefficient of variation for distribution *A* is lower than that for distribution *B*. The coefficient of variation for a distribution is the ratio of the standard deviation of the distribution to the mean, i.e.: $\hat{\sigma} / \hat{\mu}$ in the estimation sample. Thus, by adding the one further property of transfer sensitivity, one can get a stronger practical result that helps rank aggregate income inequality across distributions even when their Lorenz curves cross (once). Again, this provides a ranking of overall income inequality between distributions, and not of social welfare more generally.

Implementing this stronger dominance rule is indeed feasible in light of the above development in this paper. All it requires is some revision of the Lorenz dominance PEC of Section 5.

The practical empirical criterion (PEC) for inequality dominance can now be revised as follows:

Step 1 – Same as before. Test the joint null hypothesis of equality of the two Lorenz curve ordinate vectors (i.e., $l^b - l^a = 0$) versus the alternative hypothesis of non-equality. In this case, the test statistic is, as before,

$$(\hat{l}^b - \hat{l}^a)'\left[\hat{\Phi}^a + \hat{\Phi}^b\right]^{-1}(\hat{l}^b - \hat{l}^a)\ .$$

If the null hypothesis is not rejected, the two Lorenz curves can be said to be not statistically significantly different, and further comparison is not pursued.

Step 2 – If the null hypothesis in Step 1 is rejected and there is a single crossing of Lorenz curve ordinates, then undertake separate "t-statistic" calculations for differences on each of the individual estimated Lorenz curve ordinates. If at least one of the t-statistics has the appropriate sign and is statistically significant compared to critical values on the SMM distribution with *K-1* and infinite degrees of freedom and none of the t-statistics (if any) that has the wrong sign is statistically significant (again based on the SMM distribution), then proceed to Step 3. Otherwise, do not draw any dominance inference.

Step 3 – Compare the coefficients of variation for the two distributions. If the coefficient of variation for the distribution with the initially higher Lorenz curve ordinates ($\hat{C}^a$, corresponding to distribution *A*, say) is smaller than the coefficient of variation for the other distribution ($\hat{C}^b$)[4], then one can conclude that distribution *A* statistically dominates distribution *B*. This implies dominance for all summary inequality measures – that is, they are smaller in distribution *A* than in distribution *B* – satisfying inequality properties (i)-(v).

Note that in this version of the PEC for transfer sensitivity dominance, comparison of the coefficients of variation is done simply by inspection. A stronger version of Step 3 (and hence of the PEC) could involve a formal statistical test on $\hat{C}^b - \hat{C}^a$. Since the standard error of the sample coefficient of variation has been found to be

$$S.E.(\hat{C}) = 100C\left[\frac{1+2C^2}{2N}\right]^{1/2} \qquad (44)$$

---

[4] Note that the capital letter C for coefficient of variation here is quite different from the lower case c_i for mean income contribution in Section 6 above.

where $\hat{C}$ is expressed as a proportion (Ahn and Fessler, 2003), the estimated variance of $\hat{C}^b - \hat{C}^a$ for independent samples is

$$V\hat{a}r(\hat{C}^b - \hat{C}^a) = S.E.\left(\hat{C}^a\right)^2 + S.E.\left(\hat{C}^b\right)^2 .$$

Since $\hat{C}$ is asymptotically normally distributed (Ahn and Fessler, 2003), one can do an (asymptotic) normal test on the "t-ratio" test statistic

$$t = \frac{\hat{C}^b - \hat{C}^a}{\left[V\hat{a}r(\hat{C}^b - \hat{C}^a)\right]^{1/2}} . \tag{45}$$

And given that one is interested in a one-sided alternative test hypothesis, it makes sense to perform a one-tailed test on the standard normal where $H_1 : C^b - C^a > 0$ (i.e., distribution *A* has a smaller coefficient of variation).

9. Inequality Dominance with Multiple Lorenz Curve Crossings

But what if we have a situation where two Lorenz curves cross more than once? The Shorrocks and Foster (1987) approach has indeed been extended by Davies and Hoy (1994) to handle just this situation and can be viewed as a generalization of the former. In this case, instead of a single crossing and single coefficient of variation test, Davies and Hoy (1994) allow for possibly multiple Lorenz curve crossings – in the current author's experience two crossings is the most ever seen and the typical number of crossings is one – and posit a coefficient of variation condition for each cross-over point (including the top right-hand (1,1) point on the Lorenz curves).

More specifically, where two distributions *A* and *B* are being compared, Davies and Hoy (1994) show that the following statements are equivalent:

1) For all summary measures of inequality, *I*, satisfying inequality properties (i)-(v) – i..e., including transfer sensitivity – $I^a < I^b$; and

2) For all cross-over points *k = 1, 2 …,* the cumulative coefficients of variation at point *k* are smaller in distribution *A* than *B*.

To empirically implement this, again represent the two Lorenz curves being compared by vectors of their (sample) ordinates. Consider also what we will call cumulative or conditional coefficients of variation corresponding to each of the quantile cut-offs, $\xi_1, \ldots, \xi_{K-1}$, and for the full sample as well. In terms of notation, let the cumulative coefficients of variation be Cc_i , where

$$Cc_i^2 = E((Y - \mu c_i)^2 \mid Y \leq \xi_i) \,/\, [E(Y \mid Y \leq \xi_i)]^2 ,$$

$\mu c_i = E(Y \mid Y \leq \xi_i)$ is the cumulative mean (up to ξ_i),

and $\sigma c_i^2 = E((Y - \mu c_i)^2 \mid Y \leq \xi_i)$ is the cumulative variance.

So $Cc_i = \sigma c_i \,/\, \mu c_i$. The unconditional coefficient of variation for the full set of observations can be viewed as the case of *i*=*K* (i.e., $Cc_K = C$). Then a PEC for the Davies-Hoy situation can be stated as follows.

<u>Step 1</u> – Same as before. Test the joint null hypothesis of equality of the two Lorenz curve ordinate vectors (i.e., $l^b - l^c = 0$) versus the alternative hypothesis of non-equality. In this case, the test statistic is again

$$(\hat{l}^b - \hat{l}^a)'\left[\hat{\Phi}^a + \hat{\Phi}^b\right]^{-1}(\hat{l}^b - \hat{l}^a) .$$

If the null hypothesis is not rejected, the two Lorenz curves can be said to be not statistically significantly different, and further comparison is not pursued.

<u>Step 2</u> – Essentially the same as for the single-crossing case. If the null hypothesis in Step 1 is rejected and there are one or more crossings of Lorenz curve ordinates, then undertake

separate “t-statistic” calculations for differences on each of the individual estimated Lorenz curve ordinates. If at least one of the t-statistics has the appropriate sign and is statistically significant compared to critical values on the SMM distribution with *K-1* and infinite degrees of freedom and none of the t-statistics (if any) that has the wrong sign is statistically significant (again based on the SMM distribution), then proceed to Step 3. Otherwise, do not draw any dominance inference.

Step 3 – Compare the cumulative coefficients of variation for the two distributions. As for Step 2, undertake separate “t-statistic” calculations for differences on each of the individual estimated cumulative coefficients of variation (as well as the estimated standard coefficient of variation). If at least one of the t-statistics has the appropriate sign and is statistically significant compared once again to critical values on the SMM distribution with *K* and infinite degrees of freedom and none of the t-statistics (if any) that have the wrong sign is statistically significant (again based on the SMM distribution), then one can conclude that the distribution with the initially higher Lorenz curve ordinates (distribution *A*, say) statistically dominates distribution *B*. Once again, this implies dominance for all summary inequality measures – that is, they are smaller in distribution *A* than in distribution *B* – satisfying inequality properties (i)-(v).

To implement this PEC, then, involves doing statistical inference on the $\hat{C}c_i$'s and specifically establishing the variance structure of the set of cumulative coefficients of variation.

9.1 Variance Structure of the Cumulative Coefficients of Variation

Since the coefficient of variation is the ratio of first and second moments, it makes sense that its (cumulative) sample estimates would be asymptotically normally distributed. And since we are interested only in "t-ratios" of differences, we need focus just on the variance structure of the (cumulative) sample estimates rather than the full variance-covariance structure.

We begin by recognizing that the (cumulative) coefficients of variation are continuous differentiable function of the ξ_i's and we know the (asymptotic) distribution of the sample quantile cut-off estimates, $\hat{\xi}_i, i = 1, \ldots, K-1$. Since $\hat{\xi}_i, \ldots, \hat{\xi}_{K-1}$ are asymptotically joint normal and the Cc_i's are continuous differentiable functions of the ξ_i's, then Rao's linkage theorem says that the set of $\hat{C}c_i$'s are also asymptotically joint normally distributed. Indeed, since each Cc_i is a function of only a single ξ_i, the (asymptotic) variance of $\hat{C}c_i$ is given by simply the single derivative

$$Asy.var(\hat{C}c_i) = \left(\frac{\partial Cc_i}{\partial \xi_i}\right)^2 \bullet Asy.var(\hat{\xi}_i) \qquad \text{for } i = 1, \ldots, K\text{-}1, \qquad (46)$$

where we have already seen that

$$Asy.var(\hat{\xi}_i) = \frac{p_i(1-p_i)}{[f(\xi_i)]^2}$$

where $f(\bullet)$ is the underlying population density function of incomes from which the estimation sample is drawn.

To establish the derivative $\left(\frac{\partial Cc_i}{\partial \xi_i}\right)$, it is convenient to express the (cumulative) variance as

$$\sigma c_i^2 = E(Y^2 \mid Y \leq \xi_i) - \mu c_i^2$$

and consider the derivative of Cc_i^2 :

$$\frac{\partial Cc_i^2}{\partial \xi_i} = \frac{\partial(\sigma c_i^2 / \mu c_i^2)}{\partial \xi_i}$$

$$= \left(\frac{1}{\mu c_i^2}\right) \bullet \left[\frac{\partial \sigma c_i^2}{\partial \xi_i}\right] + \sigma c_i^2 \bullet \left[\frac{\partial \mu c_i^{-2}}{\partial \xi_i}\right]$$

i.e., $2Cc_i \bullet \left[\frac{\partial Cc_i}{\partial \xi_i}\right] = \left(\frac{1}{\mu c_i^2}\right) \bullet \left[\frac{\partial \sigma c_i^2}{\partial \xi_i}\right] - 2\left(\frac{\sigma c_i^2}{\mu c_i^2}\right)\left(\frac{1}{\mu c_i}\right) \bullet \left[\frac{\partial \mu c_i}{\partial \xi_i}\right]$.

So,

$$\frac{\partial Cc_i}{\partial \xi_i} = \left(\frac{1}{2}\right)\left(\frac{1}{\sigma c_i \bullet \mu c_i}\right) \bullet \left[\frac{\partial \sigma c_i^2}{\partial \xi_i}\right] - \left(\frac{Cc_i}{\mu c_i}\right) \bullet \left[\frac{\partial \mu c_i}{\partial \xi_i}\right]. \tag{47}$$

By Leibritz's rule, the two component derivatives are then, if $\mu c_i = \left(\frac{1}{p_i}\right) \int_0^{\xi_i} y\, f(y) dy$,

$$\frac{\partial \mu c_i}{\partial \xi_i} = \left(\frac{1}{p_i}\right) \xi_i \bullet f(\xi_i), \tag{48}$$

and if $\sigma c_i^2 = E(Y^2 \mid Y \leq \xi_i) - \mu c_i^2$

$$= \left(\frac{1}{p_i}\right) \int_0^{\xi_i} y^2 f(y) dy - \mu c_i^2,$$

$$\frac{\partial \sigma c_i^2}{\partial \xi_i} = \left(\frac{1}{p_i}\right) \xi_i^2 \bullet f(\xi_i) - 2\mu c_i \bullet \left[\left(\frac{1}{p_j}\right) \xi_i \bullet f(\xi_i)\right]$$

$$= \left(\frac{1}{p_i}\right) \xi_i \bullet f(\xi_i) \bullet [\xi_i - 2\mu c_i]. \tag{49}$$

Substituting (48) and (49) into (47) leads to

$$\frac{\partial Cc_i}{\partial \xi_i} = \left(\frac{1}{2}\right)\left(\frac{1}{\sigma c_i \bullet \mu c_i}\right) \bullet \left[\left(\frac{1}{p_i}\right) \xi_i \bullet f(\xi_i)(\xi_i - 2\mu c_i)\right]$$

$$- \left(\frac{Cc_i}{\mu c_i}\right) \bullet \left[\left(\frac{1}{p_i}\right) \xi_i \bullet f(\xi_i)\right]$$

$$= \left(\frac{1}{p_i}\right) \xi_i \bullet f(\xi_i) \bullet \left\{\left(\frac{1}{2}\right) \frac{(\xi_i - 2\mu c_i)}{\sigma c_i \bullet \mu c_i} - \left(\frac{Cc_i}{\mu c_i}\right)\right\}.$$

Thus,

$$Asy.var\left(\hat{C}c_i\right) = \left(\frac{\partial Cc_i}{\partial \xi_i}\right)^2 \bullet \frac{p_i(1 - p_i)}{[f(\xi_i)]^2}$$

$$= \left(\frac{1 - p_i}{p_i}\right)\left(\frac{\xi_i}{\mu c_i}\right)^2 \bullet \left\{\left(\frac{1}{2}\right) \frac{(\xi_i - 2\mu c_i)}{\sigma c_i} - Cc_i\right\}^2$$

$$= \left(\frac{1 - p_i}{p_i}\right)\left(\frac{\xi_i}{\mu c_i}\right)^2 \bullet \left\{\left(\frac{1}{2}\right)\left(\frac{1}{Cc_i}\right) \bullet \left[\left(\frac{\xi_i}{\mu c_i}\right) - 2\right] - Cc_i\right\}^2. \tag{50}$$

Note that, once again, the (asymptotic) variance is distribution-free, and each term in (50) can be consistently estimated from the available sample.

The standard error of the sample $\hat{C}c_i$, then, is gotten as

$$S.E.\left(\hat{C}c_i\right) = \left[\frac{Asy\hat{.}var(\hat{C}c_i)}{N}\right]^{1/2}$$

where, as usual, *N* is the size of the estimation sample. The estimated variance of the difference in (cumulative) coefficients of variation for independent samples from distribution *A* and *B* is

$$V\hat{a}r(\hat{C}c_i^{\,b} - \hat{C}c_i^{\,a}) = S.E.\left(\hat{C}c_i^{\,a}\right)^2 + S.E.\left(\hat{C}c_i^{\,b}\right)^2$$

and the t-ratio statistic of the difference is then

$$t = \frac{\hat{C}c_i^{\,b} - \hat{C}c_i^{\,a}}{\left[V\hat{a}r(\hat{C}c_i^{\,b} - \hat{C}c_i^{\,a})\right]^{1/2}}$$

that is used in Step 3 of the PEC test criterion.

The Appendix at the end of this paper shows how the $\hat{C}c_i$'s can be calculated from quantile-specific means and standard deviations by straightforward recursion formulas. So, again, it would be helpful to practitioners using official published statistics on income shares if the official statistical agencies also provided quantile standard deviations along with their quantile means, so users can undertake statistical inference on Lorenz curves if they wish.

10. Distributional Distance Dominance

One aspect of concern about rising income inequality is the implied growing economic and social distance between income groups and the potential political fracturing this may bring about. The literature and media have focussed on the widening gap between top incomes and the rest of the distribution and the increasing difficulty of lower-income workers to pull ahead into stable middle-income status – the sense of belonging to the Middle Class may be weakening. So this raises the question of whether there is a way to measure, in general fashion, the growing economic distances between different income groups across a distribution?

The analysis of this paper suggests just such a measure – a "distributional distance function". It can perhaps be most conveniently pictured as a graph with deciles or percentiles measured along the base or horizontal axis (e.g., $i = 1, 2, \ldots, K$) and incremental quantile mean income gaps $(\hat{\mu}_i - \hat{\mu}_{i-1})$ measured along the vertical or left-hand axis. This relationship or curve may be referred to as the distributional distance function for a given income distribution.

Its ordinates show the distance or income gap between adjacent quantile groups in a distribution. So for some distributions, the gaps may be relatively wide between lower and middle-class quantile groups, suggesting it is more difficult to move up to middle-income status. Obviously, gaps could be combined to show the distance between, say, bottom and middle income groups. While for other situations – such as over recent decades in the Canadian and U.S. economies – the widening gaps have been most dramatic at the upper end of the distributions.[5]

Indeed, one can compare such curves between two income distributions and argue that the uniformly lower curve is said to distance dominate or show distributional distance dominance over the higher such curve. Comparing such curves would also allow one to identify which regions of a distribution are showing widening income distance gaps over time. Such comparisons can be easily done from decile mean income figures published annually by official statistical agencies such as Statistics Canada and the U.S. Bureau of the Census.

[5] Distributional distances could also be expressed in proportion terms – such as $(\hat{\mu}_i - \hat{\mu}_{i-1}) / \hat{\mu}_{i-1}$ – and application of Rao's linkage theorem would still carry through. But for intuitive appeal and convenience of linear analysis, we'll express distributional distances in dollar or level terms.

Employing the analytical machinery of the present paper also allows one to extend formal statistical inference and hypothesis testing to such a comparison. The key is to represent each (sample) distributional distance function by the vector

$$\hat{d} = (\hat{\mu}_1,\ \hat{\mu}_2 - \hat{\mu}_1, \dots,\ \hat{\mu}_K - \hat{\mu}_{K-1})'$$

where the first element can be thought of as $\hat{\mu}_1 - \hat{\mu}_0$ where $\hat{\mu}_0 = 0$. Testing and inference then depend on the statistical properties of these quantile mean differences. One can also develop a formal PEC rule for comparing income distributions in terms of overall distance dominance.

10.1 Statistical Inference for the Distributional Distance Function

The sample vector $\hat{d}$ is a linear function of $\hat{\mu}$, the vector of quantile means:

$$\hat{d} = D \bullet \hat{\mu} \qquad \text{and } d = D \bullet \mu$$

where, for purposes of this section,[6]

$$D = \begin{bmatrix} 1 & & & 0 \\ -1 & 1 & & \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{bmatrix} \tag{51}$$

is a *KxK* non-random banded matrix with ones along the principal diagonal, minus ones just below the principal diagonal, and zeros elsewhere. As has already been seen, $\hat{\mu}$ is asymptotically joint normally distributed with mean vector μ and (asymptotic) variance-covariance matrix V_S . It then follows from Rao's linkage theorem that $\hat{d}$ is also asymptotically joint normally distributed with mean d and (asymptotic) variance-covariance matrix

$$Asy.var(\hat{d}) = \Delta_S = D \bullet V_S \bullet D' = [\delta_S(i,j)] \,, \tag{52}$$

where the elements of V_S are worked out in section 4.2 above. Thus, for example, for *i = 2, …, K*,

---

[6] This is to be distinguished from the *D* matrix in Section 7.

$$Asy.var(\hat{d}_i) = v_S(i-1, i-1) + v_S(i,i) - 2v_S(i-1, i)$$

where $v_S(i,j)$ is the *i,j*'th element of V_S . Thus, the standard error of $\hat{d}_i$ is

$$S.E.(\hat{d}_i) = \left[\frac{Asy.\widehat{var}(\hat{d}_i)}{N}\right]^{1/2} \tag{53}$$

for estimation sample size *N*. One can then use (53) to formally test for the statistical significance of any individual quantile difference based on an (asymptotic) standard normal test.

10.2 A PEC for Distributional Distance Dominance

To rank overall distance dominance between two (independent) income distributions, one can adopt a practical empirical criterion similar to that for establishing rank dominance.

Step 1 – Test the joint null hypothesis of equality of the two distributional distance vectors, d^a and d^b (corresponding to distributions *A* and *B*), versus the alternative hypothesis of non-equality. This can be done with the test statistic

$$(\hat{d}^b - \hat{d}^a)' \left[\hat{\Delta}^a + \hat{\Delta}^b\right]^{-1} (\hat{d}^b - \hat{d}^a) \tag{54}$$

where $\hat{\Delta}^a = \left[\hat{\delta}^a(i,j)\right]$, $\quad \hat{\Delta}^b = \left[\hat{\delta}^b(i,j)\right]$,

$\hat{\delta}^a(i,j) = \hat{\delta}^a_S(i,j) \,/\, N^a$,

and $\quad \hat{\delta}^b(i,j) = \hat{\delta}^b_S(i,j) \,/\, N^b$.

$\hat{\Delta}^a$ and $\hat{\Delta}^b$ are thus the estimated variance-covariance matrices of $\hat{d}^a$ and $\hat{d}^b$, respectively, and are obtained by (i) rescaling the elements of the asymptotic variance-covariance matrices Δ^a_S and Δ^b_S by their respective sample sizes and (ii) replacing all unknown population terms by their consistent sample estimates. Under the null hypothesis of equality of the two distance vectors, statistic (54) is asymptotically chi-square with *K* degrees of freedom. If the null hypothesis is not rejected, then the two distributions can be said to have distributional distance functions that are not statistically significantly different, and further comparison is not pursued.

Step 2 – If, however, the null hypothesis in Step 1 is rejected, then proceed to calculate separate "t-statistics" for differences on each of the individual quantile distance elements:

$$t_i = \frac{\hat{d}_i{}^b - \hat{d}_i{}^a}{\left[\hat{\delta}^a(i,i) + \hat{\delta}^b(i,i)\right]^{1/2}}, \qquad i = 1, \ldots K, \tag{55}$$

where $\hat{\delta}(i,i) = \left[S.E.(\hat{d}_i)\right]^2$ is the estimated variance of $\hat{d}_i$ – disregarding the superscripts *a* and *b* for convenience. Then compare these individual t-statistics to critical values on the SMM distribution with *K* and infinite degrees of freedom. If at least one of these individual "t-tests" is statistically significant of one sign and none of the other individual t-statistics are statistically significant of the other sign, then one can conclude that the distributional distance function with the lower $\hat{d}_i$'s – say distribution *A* – dominates that of the other distribution (i.e., *B*). If not, one can say that the two distributional distance functions are statistically significantly different, but not reach a conclusion as to whether one distance dominates the other overall.

11. Review and Conclusions

The theoretical literature on social choice and economic welfare evaluation has offered several dominance criteria for ranking different income distributions – such as rank dominance, Lorenz dominance or generalized Lorenz dominance – and based on comparing curves such as quantile mean curves or Lorenz curves. This paper provides the tools and procedures for actually implementing these dominance criteria empirically with data that can be readily obtained from statistical agencies such as Statistics Canada and the U.S. Bureau of the Census. The approach followed thus advances the statistical inference framework for a toolbox of disaggregative income inequality measures (such as quantile means and income shares) published by these agencies.

The process for implementing this advance involves three stages. The first stage consists of representing a dominance curve by a vector of the curve's estimated ordinates for a set of specified quantile points (such as deciles or percentiles). This transforms a theoretical problem into a statistical one. The second stage involves establishing the statistical properties of this vector of sample ordinates through use of recent developments on quantile-based inferences that are distribution-free and thus very straightforward to implement. This transforms the statistical problem into an inferential one by providing a framework for basing comparisons on formal statistical inference and testing. The third stage of implementing dominance comparisons involves proposing specific practical empirical criteria (or PECs) – one can think of these as a type of decision tree – for using formal statistical inference tests to reach empirical conclusions about the ranking of income distributions based on the theoretical dominance criteria. This converts a series of statistical test outcomes to conclusions with respect to the possible ranking of income distributions being compared.

This approach is applied to several dominance rules for ranking social welfare or income inequality between distributions:

- rank dominance for comparing social welfare (Section 4)
- Lorenz dominance for comparing income inequality (Section 5)
- generalized Lorenz dominance for comparing social welfare (Section 6)
- Lorenz dominance for comparing income inequality when Lorenz curves cross (Sections 8, 9)

- distributional distance dominance for comparing distances between income groups (Section 10).

The latter is a novel concept to highlight another feature or property of an income distribution. These dominance criteria can all be expressed in terms of linear transforms of income shares and quantile means, and hence their statistical properties can be easily established. Since statistical inference for quantile-based income shares and quantile means has been shown to be distribution-free (in the sense of not depending on any specific underlying income distribution function), so also is statistical inference for these transforms, and hence test statistics can be readily obtained as well.

The analytical results in the paper have several implications. First, they show that quite broad inferences can be drawn as to social welfare and inequality comparisons that do not rely on single specific measures and can be much more general. Thus there should be a shift in focus from specific summary measures of inequality to whole sets of disaggregative measures that are readily available in official statistical sources. Since these disaggregative measures are all quantile-based, the analysis thus highlights this disaggregative quantile-based approach to characterizing and measuring income inequality.

Second, the analysis of the paper shows that these disaggregative income inequality statistics can – when used jointly – provide not just descriptive information on changing patterns of inequality, but also (under fairly broad and reasonable conditions) *normative* insights and inferences as well. The paper also shows how these readily available disaggregative measures of income inequality – with only a bit more information that could be easily provided by statistical agencies – can provide the basis for formal *statistical inference* and standard statistical testing protocols.

Third, the set of proposed toolbox measures of disaggregative income inequality in Beach (2021a,b) should be expanded to include several further measures highlighted in the current paper:

- Lorenz curve ordinates (Section 5)
- generalized Lorenz curve ordinates (Section 6)
- relative-mean income ordinates (Section 7),
- quantile standard deviations (Section 9), and
- distributional distance measures (Section 10).

The paper provides full variance-covariance formulas for each of these measures as well, so they can all be used in informative and straightforward fashion.

Fourth, government statistical agencies such as Statistics Canada and the U.S. Bureau of the Census should provide – along with their income shares, quantile means and quantile cut-off values published annually – information on (i) the sample sizes of the estimation samples the above statistics are based on, as well as (ii) cumulative means and standard deviations by quantile group (including for the full samples). This would allow empirical users to calculate relevant test statistics for formal statistical inference on the above published statistics as part of their empirical analysis.

And fifth, the inferential approach in this paper could also be combined with active on-going research on distributional National Accounts (Zucman et al., 2018; Alvaredo et al., 2020).

References

Ahn, S., and J.A. Fessler (2003) "Standard Errors of Mean, Variance and Standard Deviation Estimators", available on-line at: web.eecs.umich.edu/~fessler/papers/files/tr/stderr.pdf

Alvaredo, F., A.B. Atkinson, L. Chancel, T. Piketty, E. Saez, and G. Zucman (2020) "Distributional National Accounts Guidelines: Methods and Concepts Used in WID.world", World Inequality Database, WID Working Paper No. 2016/2 (version of April 5, 2018).

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2018) *For Good Measure: Advancing Research on Well-being Metrics Beyond GDP*, Organization for Economic Cooperation and Development, chapter on Distributional National Accounts, 143-161.

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

Beach, C.M. (2021a) "A Useful Empirical Tool Box for Distributional Analysis", Queen's University, Department of Economics Discussion Paper No. 1466, August.

__________ (2021b) "A Nifty Fix-for Published Distribution Statistics: Simplified Distribution-Free Statistical Inference", Queen's University, Department of Economics Discussion Paper No. 1477, September.

Beach, C.M, R. Davidson, and G.A. Slotsve (1994) "Distribution-Free Statistical Inference for Inequality Dominance with Crossing Lorenz Curves", Queen's University, Department of Economics Discussion Paper No. 912, November.

Beach, C.M, and J. Richmond (1985) "Joint Confidence Intervals for Income Shares and Lorenz Curves", *International Economic Review* 26(2), 439-450.

Bishop, J.A., J.P. Formby, and P.D. Thistle (1989) "Statistical Inference, Income Distributions and Social Welfare", in *Research on Economic Inequality*, vol. 1, ed. D.G. Slottje. Greenwich, CT: JAI Press, 49-82.

__________ (1992) "Convergence of the South and Non-South Income Distribution, 1969-1979", *American Economic Review* 82(1), 262-272.

Boadway, R.W., and N. Bruce (1984) *Welfare Economics*. Oxford: Basil Blackwell Inc.

Davidson, R. (2018) "Statistical Inference on the Canadian Middle Class". *Econometrics*, Special Issue on Econometrics and Inequality 6(1), 14, 1-18.

Davies, J., and M. Hoy (1994) "Making Inequality Comparisons when Lorenz Curves Intersect", *American Economic Review* 85(3), 980-986.

Department of Finance Canada (2021) "Measuring What Matters: Toward a Quality of Life Strategy for Canada", April 19, 2021 report available on-line at: Canada.ca/en/department-finance/services/publications/measuring-what-matters-toward-quality-life-strategy-canada.html

Drummond, D. (2021) "Viewpoint: Canada Should Establish an Equitable Growth Institute", Centre for the Study of Living Standards, *International Productivity Monitor*, No. 38 (Spring).

The Economist (2021a) "Free Exchange: Fleshing out the Olive", Aug. 28, 2021, 65.

__________ (2021b) "Free Exchange: Black Cat, White Cat, Fat Cat, Thin Cat", Oct. 2, 2021, 62.

__________ (2021c) "Incomes: Piketty Lines", Oct. 9, 2021, 24.

Jenkins, S. (1991) "The Measurement of Income Inequality", in L. Osberg (ed.), *Economic Inequality and Poverty: International Perspectives*. Armonk, N.Y.: M.E. Sharpe Inc., 3-38.

Jorgenson, D.W. (1990) "Aggregate Consumer Behavior and the Measurement of Social Welfare", *Econometrica* 58(5), 1007-1040.

Kendall, M.G., and A Stuart (1969) *The Advanced Theory of Statistics*, vol. 1. London: Charles Griffen & Co.

Lin, P.E., K.T. Wu, and I.A. Ahmad (1980) "Asymptotic Joint Distribution of Sample Quantiles and Sample Mean with Applications", *Communications in Statistics – Theory and Methods* 9(1), 51-60.

Lohr, S. (2022) "Economists Eye Tech's Influence on Inequality", *The Globe and Mail* Report on Business, Jan 12, B4.

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

Saposnik, R. (1981) "Rank Dominance in Income Distribution", *Public Choice* 36(1), 147-151.

Shorrocks, A.F. (1983) "Ranking Income Distributions", *Economica* 50(197), 3-17.

Shorrocks, A.F., and J. Foster (1987) "Transfer Sensitive Inequality Measures", *Review of Economic Studies* 54, 485-497.

Stoline, M.R., and H. K. Ury (1979) "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons Among Means", *Technometrics* 21(1), 87-93.

Wilks, S.S. (1962) *Mathematical Statistics*. New York: John Wiley & Sons.

Wright, D. (2021) “Rhetoric vs Results: Shaping Policy to Benefit Canada’s Middle Class”, Public Policy Forum report, Ottawa (June).

Zucman, G., T. Piketty, and E. Saez (2018) “Distributional National Accounts: Methods and Estimates for the United States”, *Quarterly Journal of Economics* 133(2), 553-609.

Appendix

Calculating Cumulative Coefficients of Variation from Quantile Means and Standard Deviations

This appendix shows how one can calculate cumulative coefficients of variation (for use in statistical inference in Section 9 on crossing Lorenz curves) from quantile-specific means and standard deviations by simple recursion formulas.

Let the cumulative coefficient of variation

$$Cc_i \equiv \sigma c_i \,/\, \mu c_i \tag{a1}$$

where the lower-case c_i's indicate cumulative; i.e.,

$$\mu c_i = E(Y \mid Y \leq \xi_i)$$

and $$(\sigma c_i)^2 = Var(Y \mid Y \leq \xi_i)\ .$$

We develop recursion formulas for each of μc_i and $(\sigma c_i)^2$. In the case of the means,

$$E(Y \mid Y \leq \xi_i) = \left(\frac{p_{i-1}}{p_i}\right) E(Y \mid Y \leq \xi_{i-1}) + \left(\frac{D_i}{p_i}\right) E(Y \mid \xi_{i-1} < Y \leq \xi_i)$$

i.e.: $$\mu c_i = \left(\frac{p_{i-1}}{p_i}\right) \mu c_{i-1} + \left(\frac{D_i}{p_i}\right) \mu_i \tag{a2}$$

for $i = 2, \ldots, K$ (where $\xi_K = \infty$) and with $\mu c_1 = \mu_1$.

In the case of standard deviations, it is more convenient to work with the variances:

$$(\sigma c_i)^2 = E[(Y - \mu c_i)^2 \mid Y \leq \xi_i]$$

$$= E\,(Y^2 \mid Y \leq \xi_i) - (\mu c_i)^2\ . \tag{a3}$$

But the same reasoning as in (a2) holds for cumulating the Y^2's:

$$E\,(Y^2 \mid Y \leq \xi_i) = \left(\frac{p_{i-1}}{p_i}\right) E\,(Y^2 \mid Y \leq \xi_{i-1}) + \left(\frac{D_i}{p_i}\right) E(Y^2 \mid \xi_{i-1} < Y \leq \xi_i)$$

i.e.: $$\mu c2_i = \left(\frac{p_{i-1}}{p_i}\right) \mu c2_{i-1} + \left(\frac{D_i}{p_i}\right) \mu 2_i \tag{a4}$$

for $i = 2, \ldots, K$, and with $\mu c2_1 = (\sigma_1)^2 + (\mu_1)^2$, where the *2* in each term of (a4) refers to operations on the Y^2's.

Therefore, plugging recursion formulas (a2) and (a4) into equations (a3) and (a1) allows one to update $(\sigma_i)^2$ and thus Cc_i as well for each $i = 2, \ldots, K$.

Given a set of sample quantile means and standard deviations, then, one can calculate the cumulative estimated coefficients of variation in analogous fashion.

$$\mu\hat{c}_i = \left(\frac{p_{i-1}}{p_i}\right)\mu\hat{c}_{i-1} + \left(\frac{D_i}{p_i}\right)\hat{\mu}_i \qquad \text{(a5)}$$

for $i = 2, \ldots, K$ with $\mu\hat{c}_1 = \hat{\mu}_1$. Since, $\sigma_i^2 = \mu 2_i - (\mu_i)^2$,

$$\mu 2_i = \sigma_i^2 + (\mu_i)^2 \ .$$

So $\quad \mu\hat{2}_i = \hat{\sigma}_i^2 + (\hat{\mu}_i)^2 \quad$ for $i = 1, \ldots, K$ (a6)

and $\hat{\sigma}_i^2$ is the square of the *i*'th quantile sample standard deviation. Here $(\hat{\sigma}_i)^2 = (1 / N_i) \bullet \sum_{j,R_i} \left(Y_j - \hat{\mu}_i\right)^2$

where R_i refers to the range of observations greater than $\hat{\xi}_{i-1}$ and less than or equal to $\hat{\xi}_i$, and N_i is the number of observations in R_i . (Note that the denominator here is N_i rather than $N_i - 1$, so that adding up conditions hold.) Then

$$\mu\hat{c}2_i = \left(\frac{p_{i-1}}{p_i}\right)\mu\hat{c}2_{i-1} + \left(\frac{D_i}{p_i}\right)\hat{\mu}2_i \qquad \text{(a7)}$$

for $i = 2, \ldots, K$ with again $\mu\hat{c}2_1 = \hat{\mu}2_1$. Then

$$(\sigma\hat{c}_i)^2 = \mu\hat{c}2_i - (\hat{\mu}c_i)^2 \qquad \text{(a8)}$$

and thus $C\hat{c}_i = \sigma\hat{c}_i / \hat{\mu}c_i$. (a9)

Since the calculations involve sums and differences of squared terms, they should be performed in high precision. Note also that $C\hat{c}_K$ is indeed the estimated coefficient of variation for the sample as a whole.