Queen's Economics Department Working Paper No. 1466

A Useful Empirical Tool Box for Distributional Analysis

Charles Beach

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

8-2021

A Useful Empirical Tool Box for Distributional Analysis

August 2021

Charles M. Beach
Department of Economics
Queen's University
Kingston, ON, Canada

The author wishes to thank for research and technical assistance, Sam Gyetvay, Sharon Sullivan and Aidan Worswick. All errors and limitations of this paper, though, are solely the author's.

Abstract

This paper offers a tool box of disaggregative measures of distributional change, including population shares, income shares, quantile mean incomes and relative mean incomes of different income groups. It highlights median-based measures along with quintiles and deciles. It also offers formulas for the measures' standard errors and a common framework for statistical inference on these measures. Illustrating these tools with Census and LFS microdata, the paper highlights the substantial decline in earnings shares and relative mean earnings levels of middle-class workers in Canada since 1980 and the corresponding dramatic rise in these measures for higher earners in the labour market.

1. Introduction

This is a time of complex distributional changes to the economy and its workforce. In the longer run, the forces of automation are at work along with underlying trade, demographic, policy and institutional evolutions. Gathering evidence indicates that the current COVID-19 pandemic is hitting different economic, racial, social and demographic groups with very different impacts. And the various policy responses (and their changes) are having massive effects as well. These all affect different groups of the population and different regions of the income/earnings distributions in ways that can only be analyzed in quite disaggregative fashion, i.e., by breaking down the aggregate population into different groups and by examining the experiences of individuals or households across different regions of each of these group distributions (e.g., Acemoglu et al., 2016; Autor, Dorn and Hanson, 2013; Beach, 2016; Beaudry and Green, 2006; Beaudry, Green and Sand, 2016; Boudarbat, Lemieux and Riddell, 2010; DiNardo, Fortin and Lemieux, 1996; Goos, Manning and Salomons, 2014; Green, Riddell and St. Hilaire, 2016; and Green and Sand, 2015). Disaggregative measures of distributional change or detailed differences between income distributions can also serve as the basis for making social welfare judgements (or stochastic dominance evaluations) of distributional change or differences (e.g., Bishop, Formby, and Thistle, 1989, 1991, and 1992; Jenkins, 1991, Saposnik, 1981; and Shorrocks, 1983).

The principal objective of this paper is to provide an empirical tool box of statistical measures that can be used to undertake such disaggregative empirical distributional analysis. The traditional approach has been to look at quintile or decile distributional statistics provided by official statistical agencies such as Statistics Canada or the U.S. Bureau of the Census. Such measures are all percentile-based in how they disaggregate an income distribution, and estimated standard errors on these measures are not provided so that observed differences in measures across time or between groups are hard to evaluate in terms of statistical reliability. This paper seeks to extend this traditional approach in two respects: (i) by complementing the percentile measures typically used by a useful set or tool box of median-based distributional measures, and (ii) by gathering together (and in some cases deriving) the formulas for the empirical standard errors for both the percentile-based and the median-based distributional measures in one place based on a common quantile-function approach for ease of reference by applied practitioners. It is hoped that this way, conventional descriptive analysis in the area can be reoriented around a more formal statistical inference perspective. The paper also provides some empirical results using these tools so as to illustrate some of the benefits and insights from the use of these measures and the formulas provided.

The paper proceeds as follows. The next section sets out the tool box set of measures of distributional analysis that is being proposed, and provides standard error formulas for these measures. Section 3 then illustrates an empirical analysis of the distributional change in earnings in Canada over 1970-2015. Section 4 provides some concluding observations. The technical appendices set out the formal development of the proposed inference formulas based on the quantile-function approach and include some novel observations for the formulas provided in the paper.

2. Tool Box Measures of Distributional Analysis and Their Standard Errors

2.1 The Tool Box Measures Used

The sets of measures of disaggregative distributional analysis can be broken into those based on two approaches: percentile-based measures and median-based measures. The former are based on dividing up an income distribution into percentage groups with equal numbers of recipient units (individuals, workers, families, households) in each group, typically decile or quintile, where the recipient units are ordered from lowest income to highest income levels. So, for example, the bottom decile contains the lowest-income ten percent of recipient units and the top quintile contains the highest-income 20 percent in the distribution. The latter approach breaks up the distribution of ordered recipient units into groups defined relative to the median income level in the distribution. So the middle-income group or so-called Middle Class of the income distribution, for example, could consist of families whose incomes lie between 50% and 150% (or 50% to 200%) of the median family income level. Official government statistical agencies typically provide distributional statistics based only on the percentile-based approach. But recent work has noted the benefits of looking at median-based measures as well (e.g., Beach, 2016).

Within each approach, one can view there as being four types of (disaggregative) distributional statistics:

i. income shares (and population shares in the median-based approach),
ii. quantile means and income cut-offs,
iii. relative mean income ratios, and
iv. income gaps and differentials.

Income shares are the proportions of total income in the distribution being received by recipient units within a particular income group (e.g., the top decile of the distribution or by the Middle

Class in the distribution). Population shares are the proportions of recipient units lying within a specified income group. In the percentile approach, this is simply prespecified as, say, 10% or 20% of recipient units. In the median-based approach, however, the proportion needs to be calculated from the data as a separate statistic. Cut-off statistics are the income levels that separate one income group from an adjacent one. So the first decile cut-off is that income level that separates recipient units in the lowest and second decile income groups. Conditional or quantile means are the average or mean incomes of the recipient units within a given income group. So the mean middle-class income is the average of all incomes belonging to the middle-class income group. Relative mean income ratios are the ratios of quantile means to the overall mean income of the distribution. Income gaps are the differences between the quantile mean incomes of two specified income groups, and income differentials are the ratios of selected quantile mean incomes.

The two approaches are geometrically closely linked. Imagine a Lorenz curve with the cumulative share of income recipients along the horizontal or base axis and their cumulative share of total income measured vertically. The Lorenz curve is uniformly upward sloped with an ever increasing slope. The percentile approach identifies points on the horizontal axis (such as 0.2, 0.5, 0.8) and these are then mapped via the Lorenz curve into points on the vertical axis to indicate corresponding income shares.

Now draw a second curve from the origin of the derivative or slope of the Lorenz curve and measure it vertically downward. So the horizontal base axis has a graph or curve (i.e., the Lorenz curve) above it and a second graph or curve (i.e., the derivative curve) reflected below it. It turns out that the derivative of the Lorenz curve is the so-called relative-mean income curve which is the ratio of the quantile income level (corresponding to a point on the horizontal axis) divided by the mean income level of the distribution as a whole. Since the slope of the Lorenz curve is strictly positive and increasing, the relative-mean income curve is a monotonic transform of the Lorenz curve itself. The mid-point along the horizontal axis of this two-panel diagram is the median. Projecting this point downward onto the relative-mean income curve yields a corresponding point onto the downward vertical axis. The median-based approach then takes multiples of this point (e.g., 50 percent, 150 percent or 200 percent) measured vertically downward. These new ordinates then map back via the relative-mean income curve onto the horizontal axis (to yield corresponding population shares) and thence via the Lorenz curve onto the upward vertical axis (to yield corresponding income shares). The two approaches are thus completely linked via the Lorenz curve and its derivative or relative-mean income curve.

More formally, we use the following notation to represent the above measures:

PS_i - population share of the i'th income group, $i=1,\ldots,K$

IS_i - income share of the i'th income group, $i=1,\ldots,K$

ξ_i - upper income cut-off of the i'th income group, $i=1,\ldots,K-1$

μ_i - conditional or quantile mean income of the i'th income group, $i=1,\ldots,K$

$\mu_i - \mu_j$ - income gap between the (quantile) mean incomes of the i'th and j'th income groups, $i \neq j$

μ_i / μ - ratio of the i'th (quantile) mean income to the overall mean, $i=1,\ldots,K$.

Sample estimates of each of these measures will be indicated in what follows by putting superscript hats on top of each of these symbols.

Percentile-based distributional measures are typically conveniently available from official statistical agencies and far and away most frequently cited in the applied literature and in the media. However, median-based measures allow for a more insightful interpretation of empirical results as they allow one to analyze both the size and the relative prosperity of the income groups separately. Percentile-based measures, by construction, assign the size of income groups as a prespecified percent (e.g., top 10%) of all income recipients. But characterizing group size and prosperity separately allows one to see quite distinct distributional patterns. In economic terms, the size effect captures the quantity dimension of a change in the group's total income, while the prosperity effect captures the (relative) price dimension. This in turn can be used to help identify on net the relative strength of demand-side or supply-side driving factors behind observed distributional changes in income shares (Katz and Murphy, 1992; Beach, 2016).

2.2 Conventional Approach to Estimating Percentile-Based Measures' Standard Errors

Disaggregative distributional measures are typically computed from large cross-sectional microdata files based on official national government surveys or censuses of households or individuals (e.g., Survey of Consumer Income or the Labour Force Survey) with large numbers of observations. With such large surveys, often seemingly minor changes over time in such statistics can still be statistically significant and estimates provided with considerable statistical confidence. But it would be useful to be able to confirm if such estimates are indeed quite so reliable. This section (and the next two) present the statistical tools to be able to do so by providing the formulas for (asymptotic) standard errors for each of the above measures.[1] Note also that the sample estimates of all of the tool box measures discussed in this paper have been found to be asymptotically normally distributed, so that conventional statistical inference procedures can be applied to all of them to at least an approximate degree. Hence the focus of

[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 weights of the observations.

this and the next two sections will be on how standard errors can be calculated for each of these estimated measures so that such inferences can proceed.

Presentation of the standard error formulas is broken into several separate sections because the percentile-based approach and the median-based approach involve rather different statistical techniques in their derivations. As will be seen below, they also have an important difference in the nature of the formulas that result; more specifically, many of the percentile-based formulas turn out to be distribution-free in that the standard error formulas can be implemented without having to know or make an assumption about the underlying statistical distribution of the data, whereas the median-based standard error formulas do need to provide some such information in order to apply them.

Formal derivation of the standard error formulas for the percentile-based approach is provided in the Technical Appendix A which follows Beach and Davidson (1983) and Beach et al. (1994). Users are assumed to have a microdata sample of *N* observations of the income variable Y whose overall mean and variance are μ and σ^2, respectively. In terms of applying the required formulas, one needs to calculate the following sequence of terms, where *K* is the number of (equal sized) percentile groups (e.g., $K=10$ for deciles), p_i is the percentile proportion of individuals (e.g., $p_8 = .80$), ξ_j is the j'th percentile cut-off income level $(p_1, \ldots, p_{K-1})$, γ_i is the (cumulative) mean for the i'th (cumulative) percentile group (*i.e.*, $\gamma_i = E(Y/Y \leq \xi_i)$) and λ_i^2 is the (cumulative) variance for the i'th (cumulative) percentile group (i.e., $\lambda_i^2 = var\ (Y/Y \leq \xi_i)$. The quantile mean then is $\mu_i = E(Y \mid \xi_{i-1} < Y \leq \xi_i)$.

To calculate the standard errors required, first calculate sample estimates of the variance-covariance terms

$$\omega_{ij} = p_i[\lambda_i^2 + (1-p_i)(\xi_i - \gamma_i)(\xi_j - \gamma_j) + (\xi_i - \gamma_i)(\gamma_j - \gamma_i)] \qquad (1a)$$

$$for\ i \le j = 1, \dots, K\ .$$

Call the full matrix of ω_{ij} terms the matrix *Ω*.

In the case where *i* = *j*, this reduces to

$$\omega_{ii} = p_i\left[\lambda_i^2 + (1 - p_i)(\xi_i - \gamma_i)^2\right] \qquad for\ i = 1, \dots, K\ . \tag{1b}$$

For the sample quantile means $\hat{\mu}_1, \dots, \hat{\mu}_K$ themselves, then,

$$Asy.\ var\ (\hat{\mu}_1) = 100\ \omega_{11} \tag{2a}$$

$$Asy.\ var\ (\hat{\mu}_i) = 100\ (\omega_{ii} + \omega_{i-1,i-1} - 2\omega_{i,i-1}) \qquad \text{for } i=2,\dots,K \tag{2b}$$

again for *K*=*10* deciles. More generally, call the (asymptotic) variance-covariance matrix of the full set of *K* sample quantile means the matrix

$$V = \left[v_{ij}\right]$$

where $V = R\ \Omega\ R'$

and $$R = \begin{bmatrix} \frac{1}{p_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{p_K} \end{bmatrix} \bullet \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & 0 & \cdots & 0 & 0 \\ 0 & -2 & 3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & & -(K-1) & K \end{bmatrix}.$$

Thus the standard errors of the quantile means are

$$S.E.(\hat{\mu}_i) = \left[Asy\hat{.}var(\hat{\mu}_i)\ /\ N\right]^{1/2} \qquad i=1,\dots,K\ , \tag{3}$$

where all unknowns are replaced by their samples estimates.

In the case of income shares *(ISᵢ)*, one needs to make use of the full *KxK* (asymptotic) variance-covariance matrix of Lorenz curve ordinates

$\Theta_L = \left[\Theta_{ij}\right]$ where

$$\Theta_{ij} = \left(\frac{1}{\mu^2}\right)\omega_{ij} + \left(\frac{p_i\gamma_i}{\mu^2}\right)\left(\frac{p_j\gamma_j}{\mu^2}\right)\sigma^2 - \left(\frac{p_i\gamma_i}{\mu^3}\right)\omega_j, K+1 - \left(\frac{p_j\gamma_j}{\mu^3}\right)\omega_i, K+1 \tag{4}$$

where $\omega_j, K+1 = p_j\left[\lambda_j + \xi_j - \gamma_j)(\mu - \gamma_j)\right]$.

Then $Asy.var(I\hat{S}_1) = \theta_{11}$

and $Asy.var(I\hat{S}_i) = \theta_{ii} + \theta_{i-1,i-1} - 2\theta_{i,,i-1}$ for $i=2,\ldots,K$. (5)

Hence,

$$S.E.(I\hat{S}_i) = [Asy\hat{.}var(I\hat{S}_i) / N]^{1/2} , \tag{6}$$

where again all unknowns are replaced by their sample estimates. Since relative quantile means μ_i / μ are identically $K \bullet IS_i$, the standard error of estimates of relative quantile means is also given by

$$S.E.(\hat{\mu}_i / \hat{\mu}) = K \bullet S.E.(I\hat{S}_i) . \tag{7}$$

In the case of quantile mean differences $(\hat{\mu}_i - \hat{\mu}_j)$, again one needs to go back to the full (asymptotic) variance-covariance matrix *V*, so that

$$Asy.var(\hat{\mu}_i - \hat{\mu}_j) = v_{ii} + v_{jj} - 2v_{ij} . \tag{8}$$

Consequently,

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

$$= [\hat{v}_{ii} + \hat{v}_{jj} - 2\hat{v}_{ij}]^{1/2} / \sqrt{N} . \tag{9}$$

Since this percentile-based approach to calculating standard errors involves first calculating cumulative sample means and variances for each quantile group, this can also be referred to as a “cumulative moments approach”. This can be contrasted to an alternative “quantile function approach” that is developed in the next section for standard errors of median-based measures.

2.3 Median-Based Measures' Standard Errors

Derivation of the various formulas to calculate standard errors for the disaggregative median-based measures of income inequality is provided in Technical Appendix B.

The typical application of median-based measures is to compare how, say, the Middle Class is doing relative to those with high incomes or relative to the poor. Consequently, we illustrate this approach by dividing the distributions of income into three income groups – low incomes (L), middle incomes (M), and high incomes (H) – by two income cut-off levels expressed as multiples of the median income level, ξ, generally expressed as "a$\hat{\xi}$" for the lower cut-off and "$b\hat{\xi}$" as the upper cut-off (where, as before, the superscript hat indicates the sample estimate of the median):

lower incomes are less than $a\hat{\xi}$,

middle incomes are between $a\hat{\xi}$ and $b\hat{\xi}$,

and higher incomes are greater than $b\hat{\xi}$.

Typically, a is a fraction such as 0.5, while the upper cut-off has variously been posited as 1.5 or 2.0 (i.e., 150% or 200% of the median). For generality of presentation here, we will simply refer to $a\hat{\xi}$ and $b\hat{\xi}$.

Since the cut-offs are not specified in terms of percentiles of the distribution of income recipients, the proportions of the sample falling into the designated three groups are no longer given and need to be estimated as population shares, PS_i, $i = L,M,H$, along with estimates of the shares of total income going to the three groups as well, income shares IS_i, $i = L,M,H$.

Quantile-based distributional measures benefit from a different approach to characterizing standard errors of these estimated measures. This approach is based on recognizing that the formal definition of these measures in terms of integral functions of the underlying quantiles can be used to link the (asymptotic) distribution of sample estimates of the measures to the known (asymptotic) distribution of the estimated underlying quantiles. Hence this approach for calculating standard errors of quantile-based measures will be referred to as a "quantile function approach". It may be useful to begin by considering a very simple illustrative case.

It is shown in Appendix B that the asymptotic variance of the sample estimate $P\hat{S}_M$ of the population share for the middle-income group is

$$Asy.var\left(P\hat{S}_M\right) = [f(b\xi)(b) - f(a\xi)(a)]^2 \bullet [0.25 / f(\xi)^2] \qquad (10)$$

where *f(•)* is the underlying income distribution density function from which the data sample is drawn. $f(a\xi)$ and $f(b\xi)$ are then the values of this income density function evaluated at the cut-offs $a\xi$ and $b\xi$.

But the function *f(•)* and hence its evaluations are not known to the analyst. A way around this problem in order to be able to implement eq. (10) is to assume a functional form for *f(•)* that appears historically to be quite reasonable for income, wage and earnings distributions – a lognormal distribution. So the approach forwarded is to assume a lognormal form for the underlying income distribution, and use the available data sample to estimate the mean and variance of the (natural) log of incomes to determine estimates of the two key parameters of this distribution. Given these parameter estimates, one can generate an estimated $\hat{f}(\bullet)$ – as shown in Appendix B – which can then be used to evaluate the income density at the desired values $a\hat{\xi}$, $b\hat{\xi}$, and $\hat{\xi}$ to get $\hat{f}(\widehat{a\xi})$, $\hat{f}(\widehat{b\xi})$, and $\hat{f}(\hat{\xi})$. Replacing the unknowns in eq. (10) by these sample estimates then yields the *estimated* (asymptotic) variance

$$As\hat{y}.var\left(P\hat{S}_M\right) = \left[b \bullet \hat{f}\left(b\hat{\xi}\right) - a \bullet \hat{f}\left(a\hat{\xi}\right)\right]^2 \bullet \left[0.25 / \hat{f}(\hat{\xi})^2\right].$$

So the estimated standard error for $P\hat{S}_M$ is then

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

where *N* is the size of the available income data sample used in the estimation.

Similarly for the lower and higher income groups:

$$S.E.\left(P\hat{S}_L\right) = \left[a \bullet \hat{f}\left(a\hat{\xi}\right)\right] \bullet \left[0.5 / \hat{f}\left(\hat{\xi}\right)\right] / \sqrt{N}$$

and $$S.E.\left(P\hat{S}_H\right) = \left[b \bullet \hat{f}\left(b\hat{\xi}\right)\right] \bullet \left[0.5 / \hat{f}\left(\hat{\xi}\right)\right] / \sqrt{N} .$$

More formally, the derivations in Appendix B show that, for the lower, middle, and higher income shares:

$$Asy.var\left(I\hat{S}_L\right) = \left(\frac{1}{\mu}\right)^2 [a(a\xi) \bullet f(a\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right] + \left(\frac{1}{\mu}\right)^2 IS_L^2 \bullet \sigma^2 \qquad (11)$$

$$- 2\left(\frac{1}{\mu}\right)^2 [a(a\xi) \bullet f(a\xi)] \bullet IS_L \bullet \left[\frac{\xi-(0.5)\mu}{f(\xi)}\right],$$

$$Asy.var\left(I\hat{S}_M\right) = \left(\frac{1}{\mu}\right)^2 [b(b\xi) \bullet f(b\xi) - a(a\xi) \bullet f(a\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right]$$

$$+ \left(\frac{1}{\mu}\right)^2 IS_M^2 \bullet \sigma^2 \qquad (12)$$

$$- 2\left(\frac{1}{\mu}\right)^2 [b(b\xi) \bullet f(b\xi) - a(a\xi) \bullet f(a\xi)] \bullet IS_M \bullet \left[\frac{\xi-(0.5)\mu}{f(\xi)}\right],$$

and

$$Asy.var\left(I\hat{S}_H\right) = \left(\frac{1}{\mu}\right)^2 [b(b\xi) \bullet f(b\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right]$$

$$+ \left(\frac{1}{\mu}\right)^2 IS_H^2 \bullet \sigma^2 \qquad (13)$$

$$- 2\left(\frac{1}{\mu}\right)^2 [b(b\xi) \bullet f(b\xi)] \bullet IS_H \bullet \left[\frac{\xi-(0.5)\mu}{f(\xi)}\right].$$

For completeness, the formulas for population shares are:

$$Asy.var(P\hat{S}_L) = [a \bullet f(a\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right], \quad (14)$$

$$Asy.var(P\hat{S}_M) = [b \bullet f(b\xi) - a \bullet f(a\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right], \quad (15)$$

and $$Asy.var(P\hat{S}_H) = [b \bullet f(b\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right]. \quad (16)$$

Note that the common term in square brackets, $(0.5)^2 \ / \ f(\xi)^2 \ \equiv \ \theta$, is the asymptotic variance of the sample median, $\hat{\xi}$.

Thus the corresponding standard errors are

$$S.E.(\widehat{IS}_i) = \left[\frac{Asy.\hat{v}ar(\widehat{IS}_i)}{N}\right]^{1/2} \quad (17)$$

and $$S.E.(\widehat{PS}_i) = \left[\frac{Asy.\hat{v}ar(\widehat{PS}_i)}{N}\right]^{1/2} \quad (18)$$

for i = L, M, H where all the unknowns in eq. (11)-(16) are replaced by their sample estimates – as indicated by the "hat" on top of the asymptotic variances in (17) and (18).

Next consider the quantile means (μ_i) of the three income groups. The derivations in Appendix B then show that, for i = L, M, H,

$$Asy.var(\hat{\mu}_i) = \left(\frac{1}{PS_i}\right)^2 \bullet \ Asy.var(\widehat{N}_i) + \left(\frac{1}{PS_i}\right)^2 \mu_i^2 \bullet$$

$$Asy.var(\widehat{PS}_i) - \ 2\left(\frac{1}{PS_i}\right)^2 \mu_i \ \bullet Asy.cov(\widehat{N}_i, \widehat{D}_i) \quad (19)$$

where N_i refers to a numerator term and D_i refers to a denominator term in the derivations. Specifically, for i = L :

$$Asy.var(\widehat{N}_L) = [a(a\xi) \bullet f(a\xi)]^2 \ \bullet \ \theta$$

$$Asy.cov(\widehat{N}_L, \widehat{D}_L) = [a(a\xi) \bullet f(a\xi)] \ \bullet [a \bullet f(a\xi)] \bullet \theta \ ;$$

for i = M :

$$Asy.var(\widehat{N}_M) = [b(b\xi) \bullet f(b\xi) - a(a\xi) \bullet f(a\xi)]^2 \ \bullet \ \theta$$

$$Asy.cov(\widehat{N}_M, \widehat{D}_M) = [b(b\xi) \bullet f(b\xi) - a(a\xi) \bullet f(a\xi)] \bullet [b \bullet f(b\xi) - a \bullet f(a\xi)] \bullet \Theta \;;$$

and for $i = H$:

$$Asy.var(\widehat{N}_H) = [b(b\xi) \bullet f(b\xi)]^2 \bullet \Theta$$

$$Asy.cov(\widehat{N}_H, \widehat{D}_H) = [b(b\xi) \bullet f(b\xi)] \bullet [b \bullet f(b\xi)] \bullet \Theta$$

where, as before, Θ is the asymptotic variance of the sample median. Consequently, the corresponding standard errors are

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

for $i = L, M, H,$ where again all unknowns in (19) are replaced by their sample estimates.

Turn next to the case of the relative mean income ratio $RMI_i \equiv \mu_i \,/\, \mu$ for the three income groups. Again, the derivatives in Appendix B establish that

$$Asy.var(R\widehat{M}I_i) = \left(\frac{1}{PS_i}\right)^2 \bullet Asy.var(\widehat{IS}_i) + \left(\frac{RMI}{PS_i}\right)^2 \bullet$$

$$Asy.var(\widehat{PS}_i) - 2\left(\frac{RMI}{PS_i^2}\right) \bullet Asy.cov(\widehat{IS}_i, \widehat{PS}_i) \qquad (20)$$

for $i = L, M, H,$ where the (asymptotic) covariance term again differs across income groups.

for $i = L$:

$$Asy.cov(\widehat{IS}_L, \widehat{PS}_L) = \left[\left(\frac{1}{\mu}\right) a(a\xi) \bullet f(a\xi)\right] \bullet [a \bullet f(a\xi)] \bullet \Theta$$

$$- \left(\frac{1}{\mu}\right) IS_L \bullet [a \bullet f(a\xi)] \bullet \sigma_{12} \;;$$

for $i = M$:

$$Asy.cov(\widehat{IS}_M, \widehat{PS}_M) = \left(\frac{1}{\mu}\right) [b(b\xi) \bullet f(b\xi) - a(a\xi) \bullet f(a\xi)]$$

$$\bullet [b \bullet f(b\xi) - a \bullet f(a\xi)] \bullet \Theta - \left(\frac{1}{\mu}\right) IS_M \bullet [b \bullet f(b\xi) - a \bullet f(a\xi)] \bullet \sigma_{12} \;;$$

and for $i = H$:

$$Asy.cov\left(\widehat{IS}_H, \widehat{PS}_H\right) = \left[\left(\frac{1}{\mu}\right) b(b\xi) \bullet f(b\xi)\right] \bullet [b \bullet f(b\xi)\,] \bullet \Theta$$

$$+ \left(\frac{1}{\mu}\right) IS_H \bullet [b \bullet f(b\xi)] \bullet \sigma_{12}\,.$$

The σ_{12} term common in these three expressions is

$$\sigma_{12} = \frac{\xi - (0.5)\mu}{f(\xi)}$$

the (asymptotic) covariance between $\hat{\xi}$ and $\hat{\mu}$, the estimated median and means of the income distribution. It then follows that for *i* = *L, M, H,*

$$S.E.\left(R\widehat{M}I_i\right) = S.E.\left(\hat{\mu}_i \,/\, \hat{\mu}\right) = \left[\frac{Asy.\hat{v}ar\left(R\widehat{M}I_i\right)}{N}\right]^{1/2} \quad (21)$$

where again all unknown terms are replaced by their sample estimates.

The presentation of the calculations for the standard errors of quantile mean income gaps, $\hat{\mu}_i - \hat{\mu}_j$ for $i \neq j$ is somewhat tedious, but worth setting out so as to make explicit some of the terms in the formal derivation of Appendix B. There it is shown that:

$$Asy.var\left(\hat{\mu}_i - \hat{\mu}_j\right) = Asy.var(\hat{\mu}_i) - 2\,Asy.cov\left(\hat{\mu}_i, \hat{\mu}_j\right) + Asy.var\left(\hat{\mu}_j\right) \quad (22)$$

where the (asymptotic) variance terms have already been stated in eq. (19). But now

$$Asy.cov\left(\hat{\mu}_i, \hat{\mu}_j\right) = \left(\frac{1}{PS_j}\right)\left(\frac{1}{PS_i}\right) \bullet Asy.cov\left(\widehat{N}_j, \widehat{N}_i\right) - \left(\frac{1}{PS_j}\right)\left(\frac{\mu_i}{PS_i}\right) \bullet Asy.cov\left(\widehat{N}_j, \widehat{D}_i\right)$$

$$- \left(\frac{1}{PS_i}\right)\left(\frac{\mu_j}{PS_j}\right) \bullet Asy.cov\left(\widehat{D}_j, \widehat{N}_i\right) + \left(\frac{\mu_j}{PS_j}\right)\left(\frac{\mu_i}{PS_i}\right) \bullet Asy.cov\left(\widehat{D}_j, \widehat{D}_i\right)$$

where the various specific income-group terms need to be set out explicitly (from eq. (b15) in Appendix B):

$$Asy.cov\left(\widehat{N}_L, \widehat{N}_M\right) = g_1' g_3'\, \Theta$$

$$Asy.cov\left(\widehat{N}_L, \widehat{D}_M\right) = g_1' g_4'\, \Theta$$

$$Asy.cov\left(\widehat{D}_L, \widehat{N}_M\right) = g_2' g_3'\, \Theta$$

$$Asy.cov(\hat{D}_L, \hat{D}_M) = g_2' g_4' \, \theta$$

$$Asy.cov(\hat{N}_M, \hat{N}_H) = g_3' g_5' \, \theta$$

$$Asy.cov(\hat{N}_M, \hat{D}_H) = g_3' g_6' \, \theta$$

$$Asy.cov(\hat{D}_M, \hat{N}_H) = g_4' g_5' \, \theta$$

$$Asy.cov(\hat{D}_M, \hat{D}_H) = g_4' g_6' \, \theta$$

$$Asy.cov(\hat{N}_L, \hat{N}_H) = g_1' g_5' \, \theta$$

$$Asy.cov(\hat{N}_L, \hat{D}_H) = g_1' g_6' \, \theta$$

$$Asy.cov(\hat{D}_L, \hat{N}_H) = g_2' g_5' \, \theta$$

and $$Asy.cov(\hat{D}_L, \hat{D}_H) = g_2' g_6' \, \theta$$

where again θ is the (asymptotic) variance of the median $\hat{\xi}$, and the g_i' terms are:

$$g_1' = a(a\xi) \bullet f(a\xi)$$

$$g_2' = a \bullet f(a\xi)$$

$$g_3' = b(b\xi) \bullet f(b\xi) - a(a\xi) \bullet f(a\xi)$$

$$g_4' = b \bullet f(b\xi) - a \bullet f(a\xi)$$

$$g_5' = -b(b\xi) \bullet f(b\xi)$$

and $$g_6' = -b \bullet f(b\xi) \, .$$

As a result, then,

$$S.E.(\hat{\mu}_i - \hat{\mu}_j) = \left[\frac{Asy.\hat{var}(\hat{\mu}_i - \hat{\mu}_j)}{N}\right]^{1/2} . \qquad (23)$$

In the case of the relative or proportional mean income differential,

$$\hat{q} = \frac{\hat{\mu}_i - \hat{\mu}_j}{\hat{\mu}_j} = \frac{\hat{\mu}_i}{\hat{\mu}_j} - 1 \qquad \text{for } i > j \, ,$$

$$Asy.var(\hat{q}) = \left(\frac{1}{\mu_j}\right)^2 \bullet Asy.var(\hat{\mu}_i) + \left(\frac{\mu_i}{\mu_j^2}\right)^2 \bullet Asy.var(\hat{\mu}_j) \qquad (24)$$

$$- 2\left(\frac{1}{\mu_j}\right)\left(\frac{\mu_i}{\mu_j^2}\right) \bullet Asy.cov\left(\hat{\mu}_i, \hat{\mu}_j\right) ,$$

so that again

$$S.E.(\hat{q}) = \left[\frac{Asy\hat{.}var(\hat{q})}{N}\right]^{1/2} . \tag{25}$$

The development in Appendix B also allows one to calculate standard errors for measures of polarization and skewness of the income distribution. In the case of the polarization measure

$$P\hat{O}L = \widehat{PS}_L + \widehat{PS}_H ,$$

$$Asy.var\left(P\hat{O}L\right) = Asy.var\left(\widehat{PS}_L\right) + Asy.var\left(\widehat{PS}_H\right) + 2\, Asy.cov\left(\widehat{PS}_L , \widehat{PS}_H\right) \tag{26}$$

where the first two terms were provided in eq. (14) and (16) above and

$$Asy.cov\left(\widehat{PS}_L , \widehat{PS}_H\right) = -[a \bullet f(a\xi)] \bullet [b \bullet f(b\xi)] \bullet \theta .$$

It then follows again that

$$S.E.\left(P\hat{O}L\right) = \left[\frac{Asy\hat{.}var(P\hat{O}L)}{N}\right]^{1/2} . \tag{27}$$

A simple measure of the degree of skewness or non-symmetry in the income distribution is the ratio of the sample median to the mean, $\hat{\xi} \,/\, \hat{\mu}$. If the distribution were perfectly symmetric, the two would be exactly the same. Thus the degree to which the mean is pulled up by the distribution's long right-hand tail is picked up by the degree to which this sample ratio falls below one. A formula for the asymptotic variance of this sample ratio is provided in Lin, Wu and Ahmad (1980):

$$Asy.var\left(\hat{\xi} \,/\, \hat{\mu}\right) = \left[\left(\frac{\xi}{\mu}\right)^2 \sigma^2 - 2\left(\frac{\xi}{\mu}\right) \sigma_{12} + \sigma_{11}\right] / \ \mu^2 \tag{28}$$

where $\sigma_{11} = Asy.var\left(\hat{\xi}\right) = \frac{(0.5)(0.5)}{f(\xi)^2} = \theta$

$$\sigma_{12} = Asy.cov\left(\hat{\xi},\ \hat{\mu}\right) = \frac{[\xi - \mu(1-0.5)]}{f(\xi)}$$

and $\sigma^2 = Asy.var(\hat{\mu})$.

Consequently,

$$S.E.(\hat{\xi} \,/\, \hat{\mu}) = \left[\frac{Asy.\hat{v}ar\,(\hat{\xi} \,/\, \hat{\mu})}{N}\right]^{1/2} \tag{29}$$

where sample estimates replace all the unknowns.

2.4 A Blended Approach to Estimating Standard Errors of Percentile Cut-Offs and Percentile-Based Statistics

Another approach to calculating standard errors of percentile cut-offs and percentile-based distributional statistics involves blending aspects of the percentile-based formulation of section 2.2 and the quantile function approach of section 2.3. The result is a simplified set of formulas for such percentile-based statistics. Formal development of this blended approach and the resulting standard error formulas are found in the Technical Appendix C.

Using similar notations to section 2.2, let there be *K* (ordered) percentile income groups indexed by $i=1, \ldots, K$. Let p_i represent the (cumulative) proportions of the ordered income groups. For deciles, $p_1 = 0.1$, $p_2 = 0.2$, ..., $p_{10} = 1.0$. And again, let the income cut-offs or deciles that divide the different income groups be $\xi_1, \xi_2, \ldots, \xi_{K-1}$, and let $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \ldots \hat{\xi}_{K-1})'$ be a vector of *K-1* sample quantile cut-off estimates. Then it is well established in the statistics literature (e.g., Wilks, 1962, p. 274) that vector $\hat{\xi}$ drawn from a random sample of size *N* is asymptotically joint normally distributed with an (asymptotic) variance-covariance matrix Λ with the elements

$$Asy.var(\hat{\xi}_i) = \lambda_{ii} = p_i(1 - p_i) \,/\, f_i^2 \tag{30}$$

and $\quad Asy.cov(\hat{\xi}_i, \hat{\xi}_j) = \lambda_{ij} = p_i(1 - p_j) \,/\, f_i \bullet f_j = \lambda_{ji} \quad$ for $i < j$

where $f_i \equiv f(\xi_i) > 0$ for the underlying income distribution density function $f(\bullet)$. Clearly, these formulas are not distribution-free as the denominators involve the $f(\bullet)$ function.

But if one uses the quantile function approach based on a lognormal density of the previous sections, one can calculate $\hat{f}(\hat{\xi}_i)$ for any quantile ξ_i based on estimated lognormal distributions. Hence,

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

$$= \left[p_i(1 - p_i) \,/\, N \bullet \hat{f}(\hat{\xi}_i)^2\right]^{1/2}. \tag{31}$$

One can similarly use this approach to calculate standard errors for percentile income gaps

$\xi_j - \xi_i$ for $i \neq j$

as

$$Asy.var(\hat{\xi}_j - \hat{\xi}_i) = Asy.var(\hat{\xi}_i) - 2Asy.cov(\hat{\xi}_i, \hat{\xi}_j) + Asy.var(\hat{\xi}_j)\,, \tag{32}$$

so that

$$S.E.(\hat{\xi}_j - \hat{\xi}_i) = \left[\frac{Asy\hat{.}var(\hat{\xi}_j - \hat{\xi}_i)}{N}\right]^{1/2} \tag{33}$$

where again each of the density terms in the denominators of eq. (32) is estimated by the lognormal density.

Similarly, in the case of relative or proportional income differential

$q = (\xi_j - \xi_i) \,/\, \xi_i$ for $i < j$,

$$Asy.var(\hat{q}) = \left(\frac{\xi_j}{\xi_i^2}\right)^2 \bullet Asy.var(\hat{\xi}_i) - 2\left(\frac{1}{\xi_i}\right)\left(\frac{\xi_j}{\xi_i^2}\right) \bullet Asy.var(\hat{\xi}_i, \hat{\xi}_j)$$

$$+ \left(\frac{1}{\xi_i}\right)^2 \bullet Asy.var(\hat{\xi}_j)\,. \tag{34}$$

Consequently,

$$S.E.(\hat{q}) = \left[\frac{Asy\hat{.}var(\hat{q})}{N}\right]^{1/2} \quad (35)$$

where all unknowns in eq. (36) are estimated by the lognormal density.

For income shares based on the percentile approach, now divide the income distribution into three income groups: *L, M*, and *H*. And now let the *L* group consist of the lower $100\ p_a$ percent, the top $100(1\text{-}p_b)$ percent or H group, and the middle $100(p_b - p_a)$ percent as the *M* group where $p_b > p_a$. In empirical work, it is often taken as the bottom 20 percent, the top 20 percent, and the middle 60 percent of income recipients. In this case, $p_a = 0.2$ and $p_b = 0.8$. The corresponding percentile cut-off between the *L* and *M* income groups is ξ_a and that between the *M* and *H* groups is ξ_b .

Since percentage shares are prespecified in the percentile approach, population shares are given and have no standard errors.

Consider now the quantile mean incomes, $\hat{\mu}_L$, $\hat{\mu}_M$, and $\hat{\mu}_H$, of the three income groups. The development in Appendix C shows that their asymptotic variances, in this case, turn out to be surprisingly simple:

$$Asy.var(\hat{\mu}_L) = \left(\frac{1-p_a}{p_a}\right) \bullet \xi_a^2 \quad (36)$$

$$Asy.var(\hat{\mu}_H) = \left(\frac{p_b}{1-p_b}\right) \bullet \xi_b^2 \quad (37)$$

and

$$Asy.var(\hat{\mu}_M) = \left(\frac{1}{p_b - p_a}\right)^2 \bullet [p_a(1-p_a) \bullet \xi_a^2 + p_b(1-p_b) \bullet \xi_b^2 - 2\,p_a(1-p_b) \bullet \xi_a\xi_b]\,. \quad (38)$$

Consequently, their standard errors are

$$S.E.(\hat{\mu}_i) = \left[\frac{Asy\hat{.}var(\hat{\mu}_i)}{N}\right]^{1/2} \tag{39}$$

for $i = L, M, H$.

For the estimated relative mean income ratios, $R\widehat{M}I_i = \hat{\mu}_i \,/\, \hat{\mu}$, it can also be derived that:

$$Asy.var(R\widehat{M}I_L) = \left(\frac{1-p_a}{p_a}\right)\left(\frac{\xi_a}{\mu}\right)^2 + \frac{RMI_L^2}{\mu^2} \bullet \sigma^2 \tag{40}$$

$$+\, 2\left[\left(\frac{1-p_a}{p_a}\right)\left(\frac{\xi_a}{\mu}\right) \bullet\ RMI_L - \left(\frac{\xi_a}{\mu}\right)^2 \bullet \frac{RMI_L}{p_a}\right],$$

$$Asy.var(R\widehat{M}I_H) = \left(\frac{p_b}{1-p_b}\right)\left(\frac{\xi_b}{\mu}\right)^2 + \frac{RMI_H^2}{\mu^2} \bullet \sigma^2 \tag{41}$$

$$+\, 2\left[\left(\frac{\xi_b}{\mu}\right)^2 \bullet \frac{RMI_H}{(1-p_b)} - \left(\frac{\xi_b}{\mu}\right) \bullet\ RMI_H\right],$$

and

$$Asy.var(R\widehat{M}I_M) = \left(\frac{1}{D_M^2}\right)\left[p_a(1-\ p_a)\left(\frac{\xi_a}{\mu}\right)^2\right] + \left(\frac{1}{D_M^2}\right)\left[p_b(1-\ p_b)\left(\frac{\xi_b}{\mu}\right)^2\right]$$

$$+ \left(\frac{RMI_M}{\mu}\right)^2 \bullet \sigma^2$$

$$- 2\left(\frac{1}{D_M^2}\right)\left[p_a(1-\ p_b)\left(\frac{\xi_a}{\mu}\right)\left(\frac{\xi_b}{\mu}\right)\right] \tag{42}$$

$$+ 2\left(\frac{1}{D_M}\right)\left(\frac{\xi_a}{\mu}\right)\left(\frac{RMI_M}{\mu}\right)[\xi_a -\ \mu(1-\ p_a)]$$

$$- 2\left(\frac{1}{D_M}\right)\left(\frac{\xi_b}{\mu}\right)\left(\frac{RMI_M}{\mu}\right)[\xi_b -\ \mu(1-\ p_b)]$$

where $D_M = p_b - p_a$.

Therefore, again,

$$S.E.\left(R\widehat{M}I_i\right) = \left[\frac{Asy\hat{.}var(R\widehat{M}I_i)}{N}\right]^{1/2}. \tag{43}$$

In the case of the income shares, IS_i , since by definition of the various tool box measures

$$RMI_i \equiv \frac{\mu_i}{\mu} = \frac{IS_i}{PS_i},$$

then $IS_i = PS_i \bullet (\mu_i / \mu) = D_i \bullet RMI_i$

where the given population shares are also referred to as the denominator term D_i .

Consequently, it follows immediate that

$$Asy.var(I\hat{S}_i) = D_i^2 \bullet Asy.var(R\widehat{M}I_i) \quad (44)$$

and hence

$$S.E.(I\hat{S}_i) = D_i \bullet S.E.(R\widehat{M}I_i) \quad (45)$$

for $i = L, M, H$, so that $D_L = p_a$, $D_M = p_b - p_a$, and $D_H = 1 - p_b$.

Next consider the quantile mean income gaps, $\hat{\mu}_i - \hat{\mu}_j$ for $i \neq j$. Since the gap is the difference between two quantile means, it follows that

$$Asy.var(\hat{\mu}_i - \hat{\mu}_j) = Asy.var(\hat{\mu}_i) - 2\, Asy.cov(\hat{\mu}_i, \hat{\mu}_j) + Asy.var(\hat{\mu}_j) . \quad (46)$$

The (asymptotic) variance terms are already provided in eq. (36)-(38). For the (asymptotic) covariances, Appendix C shows that

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_M) = \frac{\xi_a}{(p_b - p_a)} [(1 - p_a)\xi_b - (1 - p_a)\xi_a] , \quad (47)$$

$$Asy.cov(\hat{\mu}_M, \hat{\mu}_H) = \frac{\xi_b}{(p_b - p_a)} [p_a\, \xi_a - p_b\, \xi_b] , \quad (48)$$

and

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_H) = -\xi_a\, \xi_b . \quad (49)$$

It thus follows that

$$S.E.(\hat{\mu}_i - \hat{\mu}_j) = \left[\frac{Asy\hat{.}var(\hat{\mu}_i - \hat{\mu}_j)}{N}\right]^{1/2} \quad (50)$$

where again all unknowns are replaced by their sample estimates.

Similarly, if one calculates the gap in relative or proportional terms as the differential

$$\hat{q} = \frac{\hat{\mu}_i - \hat{\mu}_j}{\hat{\mu}_j} = \frac{\hat{\mu}_i}{\hat{\mu}_j} = -1 \qquad \text{for } i > j ,$$

it can be seen that

$$Asy.var(\hat{q}) = \left(\frac{1}{\mu_j}\right)^2 \bullet Asy.var(\hat{\mu}_i) + \left(\frac{\mu_i}{\mu_j^2}\right)^2 \bullet Asy.var(\hat{\mu}_j)$$

$$- 2\left(\frac{1}{\mu_j}\right)\left(\frac{\mu_i}{\mu_j^2}\right) \bullet Asy.cov(\hat{\mu}_i, \hat{\mu}_j) . \qquad (51)$$

So once again,

$$S.E.(\hat{q}) = \left[\frac{Asy\hat{.}var(\hat{q})}{N}\right]^{1/2} . \qquad (52)$$

Note, incidentally that, for all of the quantile means $(\hat{\mu}_i)$, relative mean ratios $(R\widehat{M}I_i)$, income shares $(I\hat{S}_i)$ and quantile income gaps, their standard errors do not depend on evaluating a specified $f(\bullet)$ function, so statistical inference for these measures is distribution-free. This contrasts to the case of median-based measures in the previous section. Note also that the standard error formulas for the above measures are generally simpler to calculate here than in the previous section. This is because in section 2.3 the population shares need to be estimated and have to be treated as random variables as well, whereas in this section they are given as non-random D_i's expressed in terms of p_a and p_b .

Finally, work by Lin, Wu and Ahmad (1980) also allows us to undertake statistical inferences on relative mean quantile ratios, $\hat{\xi}_i \,/\, \hat{\mu}$. More specifically, they show that $\hat{\xi}_a \,/\, \hat{\mu}$ and $\hat{\xi}_b \,/\, \hat{\mu}$ are asymptotically normally distributed with (asymptotic) variances:

$$Asy.var(\hat{\xi}_a / \hat{\mu}) = \left[\left(\frac{\xi_a}{\mu}\right)^2 \sigma^2 - 2\left(\frac{\xi_a}{\mu}\right) \sigma_{13} + \sigma_{11}\right] / \mu^2 \qquad (53)$$

and

$$Asy.var(\hat{\xi}_b / \hat{\mu}) = \left[\left(\frac{\xi_b}{\mu}\right)^2 \sigma^2 - 2\left(\frac{\xi_b}{\mu}\right) \sigma_{23} + \sigma_{22}\right] / \mu^2 \qquad (54)$$

where $\sigma_{11} = \frac{p_a(1-p_a)}{f(\xi_a)^2}$ $\qquad \sigma_{22} = \frac{p_b(1-p_b)}{f(\xi_b)^2}$

$\sigma_{13} = \frac{\xi_a - \mu(1-p_a)}{f(\xi_a)}$ $\qquad \sigma_{23} = \frac{\xi_b - \mu(1-p_b)}{f(\xi_b)}$.

It then follows that

$$S.E.(\hat{\xi}_i / \hat{\mu}) = \left[\frac{A\hat{sy.var}\,(\hat{\xi}_i / \hat{\mu})}{N}\right]^{1/2} \tag{55}$$

for $i = a, b,$ where again all unknowns are replaced by their sample estimates.

Note that these standard error formulas, however, are definitely not distribution-free as the σ_{ij} terms all involve density ordinates in their denominators.

3. Illustrative Empirical Analysis of Distributional Change in Earnings in Canada, 1970-2015

3.1 Basic Data Sources and Sample Groups

The data used for this study come from Canadian Census Public Use Microdata Files (PUMF) for Individuals for 1971, 1981, 1991, 2001, and 2006, and from monthly Labour Force Survey (LFS) microdata files (for May) for each year over 2000-2015. The variable of interest is individual worker's earnings. In the Census files, earnings refers to total wage and salary income plus net self-employment income in the previous calendar year. In the LFS files, earnings refers to usual weekly wage and salary income of paid employees who are not currently full-time students. The latter thus excludes net self-employment income and the former aggregates earnings over a full year. The Census samples also refer to full-time full-year workers aged 25-59 in the earnings year, while the LFS samples refer to full-time workers (also aged 25-59) in the

LFS survey week. They are collectively referred to as full-time workers. The empirical analysis also treats male and female workers separately.

Workers in the labour market are allocated into three earnings groups – referred to as lower earners (*L*), middle-class earners (*M*), and higher earners (*H*) on the basis of their sex-specific earnings. In the case of median-based statistics:

Lower earners – those with earnings below 50% of the median (i.e., $a = 0.5$)

Middle earners – those with earnings between 50% and 150% of the median (i.e., $a = 0.5$, $b = 1.5$)

Higher earners – those with earnings above 200% of the median (i.e., $c = 2.0$).

Illustrative cut-off values – all in 2015 dollars – for the three median-based earnings groups (in earnings per week) in May 2015 are:

	Males	Females
Lower	$553	$441
Middle	$533 - $1658	$441 - $1323
Higher	$2211	$1764
Median	$1105.4	$881.8

More detailed summary statistics for the analysis samples of this study appear in Appendix Tables D1-D2.

In terms of annual figures, these cut-offs for mid 2015 correspond to:

	Males	Females
Lower	$28,741	$22,927
Middle	$28,741 - $86,221	$22,927 - $68,781
Higher	$114,962	$91,707
Median	$57,481	$45,854

At a modal hours worked per week of 37.5, these cut-offs for full-time workers also correspond to hourly cut-off values of:

	Males	Females
Lower	\$14.74	\$11.76
Middle	\$14.74 - \$44.22	\$11.76 - \$35.27
Higher	\$58.96	\$47.02
Median	\$29.48	\$23.51

The paper also looks at percentile-based distribution statistics. In this case:

Lower earners – those with earnings in the lower 20% of workers (i.e., $a = 0.2$)

Middle earners – those with earnings in the middle 60% of workers (i.e., $a = 0.2$, $b = 0.8$)

Higher earners – those with earnings in the top 20% of workers (i.e., $b = 0.8$).

Illustrative cut-off values – again in 2015 dollars – for the three percentile-based earnings groups for full-time workers (again in earnings per week) in May 2015 are:

	Males	Females
Lower	\$719	\$599
Middle	\$719 - \$1633	\$599 - \$1371
Higher	\$1633	\$1371

More detailed summary percentile-based statistics for the samples used in the study are found in Appendix Tables D3-D4. Evidently, since the lower group cut-offs for the percentile-based approach are considerably larger than for the median-based approach, the latter group includes a much smaller proportion of workers than the former. Similarly, the higher group cut-offs for the percentile-based approach are substantially smaller than for the median-based approach, so the latter group again includes a smaller proportion of workers than the former. That is to say, the *L* and *H* groups are relatively smaller – less than 20 percent each – (and hence more extreme) in the median-based approach than for the percentile-based statistics.

In terms of annual earnings figures, the percentile cut-offs correspond to:

	Males	Females
Lower	$37,398	$31,164
Middle	$37,398 - $84,900	$31,164 - $71,276
Higher	$84,900	$71,276

And based on the modal hours worked of 37.5 hours per week, the percentile cut-offs correspond to hourly cut-off values of:

	Males	Females
Lower	$19.18	$15.98
Middle	$19.18 - $43.54	$15.98 - $36.55
Higher	$43.54	$36.55

3.2 Changes in Earnings Distributional Shares in Canada

We begin by looking at the significance – both economic and statistical – of changes in distributional shares for workers' earnings in Canada over the period 1970-2015 using both Census and Labour Force Survey data and looking at both median-based and percentile-based approaches to breaking up the distribution into earnings groups. Consider first population shares from median-based estimates. Table 1 shows the percentages of (full-time) workers, separately for men and women, in the lower earnings, middle earnings, and higher earnings groups by decade over the 1970-2005 period from the Census data. Table 2 shows similar results over 5-year intervals over 2000-2015 from the LFS data. Figures in parentheses are (estimated) standard errors as specified in Appendix B of this paper. Figures in square brackets are absolute values of (asymptotic) "t-ratios" of the estimated changes in population shares.

The principal story both tables tell is that there has been a very substantial decline in the fraction of middle earners or middle-class workers in the Canadian labour market and corresponding increases in the proportions of lower-earning and higher-earning workers. This growing polarization in the labour market has been especially marked since 1980, and these changes are highly statistically significant. While this pattern holds for both male and female full-time workers, since 2005 the lower-earnings share of female workers has stopped increasing and reversed direction to start declining. The evidence is consistent with a declining proportion of middle-class workers in Canada.

Tables 3 and 4 report earnings shares of the three earnings groups estimated with the same median-based cut-offs. The tables are organized the same as for the first two tables. What results in Tables 3 and 4 show is that there has been an even larger more dramatic shift in earnings shares away from middle-class workers and to higher earners in the Canadian labour market. This is shown for both Census and LFS data sets, and again this is especially marked since 1980. Lower earnings shares also increased up until 2005 (consistent with formerly middle-class workers slipping down the earnings distributions), but since then have shown more mixed pattern of changes. There has thus been a markedly widening gap between top earnings levels and the rest of the workers in the labour market.

Why such dramatic changes have been occurring since about 1980 in the labour market – not just in Canada, but also and even moreso in the United States and in most developed economies – has produced a vast literature of contributing explanations and hypotheses. These have been extensively review in Beach (2016) and elsewhere, so will not be discussed here where the focus of the present paper is on statistical characterization of such changes. Suffice it to say here that the leading explanations for the above observed patterns of distributional change involve (i) automation and technological change arising from chip-based information and computer technology developments; (ii) increased globalization and the off-shoring of much goods production (especially in manufacturing) to areas of the world with cheaper labour costs; and (iii) changes in various institutional and policy-related matters such as deregulation in several major sectors, tax regulations affecting compensation packages, mid-1990s cut-backs in generosity and access to EI and social assistance, antiquated labour and workplace regulations, and free-trade agreements (in the U.S., declining private-sector unionization rates and falling real minimum wage rates also contributed).

The empirical results so far have been based on the median-based approach to characterizing distributional change. But as the development in Section 2 shows, an alternative percentile-based approach can also be followed, and indeed is more commonly used in published official data series (e.g., reporting series on quintile or decile income shares). We implement this here with cut-offs at the 20th percentile and 80th percentile levels. Since the percentile shares of workers are given in this approach, no population share tables (corresponding to Tables 1 and 2 above) are needed.

Earnings shares for the three percentile-based earnings groups, however, are presented in Tables 5 and 6. Again, the tables are organized the same as before. The first thing to note from these tables is that the estimated earnings shares are quite different in magnitude from those in Tables 3 and 4. The lower and higher earnings shares are much larger, consistent with many more workers being placed in these much broader categories of workers. As a result, the middle-class earnings shares are correspondingly smaller.

The second thing to note (when comparing Tables 5 and 3) is that the percentile-based results show the same pattern of falling middle-class earnings shares and rising higher earnings shares as above, but the results are more muted – though the changes are still highly statistically significant. This muting of the changes in the middle-class and higher earnings shares also shows up in Table 6 vs Table 4 for the 2000-2015 time period.

The third point of note is that changes in the lower earnings shares over the 1980-2005 period are much stronger and more dramatic in Table 5 than in Table 3 and indeed show a highly statistically significant decline in the lower earnings shares. Since 2005, however, the changes in the lower earnings shares are mixed and often not significant in Table 6.

The conclusions to be drawn are (i) the marked decline in the middle-class earnings share and corresponding dramatic rise in the higher earnings share appear quite robust to how these are measured – whether median-based or percentile-based – and (ii) what has happened to the lower earnings share, however, does appear to depend on the specifics of how the shares are estimated.

3.3 Changes in Earnings Levels and Relative Earnings by Earnings Group

The rather dramatic changes in earnings shares that have been revealed begs the question of what has actually been happening to actual earnings levels over the different regions of the earnings distribution. This can usefully be looked at in terms of conditional mean earnings, i.e., the mean earnings levels of workers in each of the three broad earnings groups (μ_i , $i = L, M, H$, in Section 2 above). Results on these conditional or quantile mean (real) earnings levels are presented in Tables 7 and 8 for the median-based estimates and in Tables 9 and 10 for the percentile-based estimates. The layouts of the tables are again the same.

The two sets of tables show virtually the same broad results. First, for male workers over the 1980s and 1990s, actual *decrease* in real (CPI adjusted) earnings levels occurred over the lower and middle-class regions of the earnings distribution and these were quite substantial:

	Median-Based (Table 7)		Percentile-Based (Table 9)	
	Lower	Middle	Lower	Middle
1980-1990	-$1664. [8.38]	-$1547. [10.03]	-$3586. [9.37]	-$1189. [4.47]
1990-2000	-$82. [0.47]	-$1718. [11.23]	-$2327. [7.82]	-$739. [3.25]

Figures in brackets are (absolute values of asymptotic) "t-ratios" of changes in indicated real earnings levels. As can be seen they are almost all highly statistically significant. For ten-year interval changes, these are historically highly unusual. In terms of percentile-based estimates, these real earnings declines have indeed continued on to 2005 as well.

Second, since 2005, real earnings increases have occurred and been fairly broadly shared across all earnings groups and for both male and female workers, though the increases have occurred at a much reduced rate than seen in the 1970s. This holds for both median-based and percentile-based results.

It was a common saying in the 1960s and 1970s that "a rising tide raises all boats". Is this true over more recent decades? To address this, we look next at relative-mean income ratios (i.e., RMI_i, $i = L, M, H$ in Section 2) or, in the present paper, the ratios of conditional mean earnings levels to the mean earnings of the earnings distribution as a whole. Results on the relative-mean earnings ratios are presented in Tables 11 and 12 for the median-based estimates and in Tables 13 and 14 for the percentile-based estimates.

First off, one sees that the relative-mean earnings (RME) ratios over the 1980-2005 interval markedly declined for lower-earnings and middle-class earnings workers, while very strongly rising for higher-earnings workers. This general result holds for both male and female workers and for both median-based and percentile-based sets of estimates (though the RME_L declines are larger when measured by the percentile approach). These changes are also highly statistically significant in all cases. The most consistent pattern of change since 1980 has been the quite marked decline in relative mean earnings for middle-class workers, which has continued since 2005 and remains highly statistically significant. Quite evidently, since the 1980s, a rising tide of generally increasing real earnings levels for the economy over the period as a whole has *not* lifted all levels evenly. Middle-class workers, especially male workers, have lost out relatively to top earners in the Canadian labour market.

3.4 Changes to the Female-to-Male Earnings Gap by Earnings Group

It is evident from the last section that there have been very marked differences between male and female earnings levels and changes in earnings levels over the entire period covered, but especially since the 1980s when women entered the Canadian full-time workforce in much greater numbers. Full-time earnings levels for women are (still) much lower than those for men, but they have generally been rising faster in percentage terms than men's full-time earnings levels.

These two patterns are highlighted in Table 15 which shows, at an aggregate level, what has happened to the female-to-male full-time earnings ratio in the Canadian labour market overall since 1970. The first column looks at relative mean earnings levels, while the second column shows the relative medians. In 1970, the female-to-male full-time earnings ratios were only 59-62 percent. By 2015, the ratios had moved up substantially to 80-82 percent (while the female participation rate had risen markedly as well). This has been quite a change in the labour market and a vast literature has been devoted to this major development. In this paper, we'll focus just on the statistics themselves. Note also that, over a 45-year interval, the full-time earnings gap has narrowed by only about half. These summary results also beg the questions of whether these relative earnings increases have been broadly experienced across the earnings distribution, or have they been driven by certain pockets or subgroups of workers.

Tables 16 and 17 address this issue. Both tables report ratios of conditional or quantile mean earnings between women and men in the full-time workforce broken down by the three earnings groups *L, M,* and *H*. The first table refers to median-based estimates and the second to percentile-based estimates. Both tables show essentially the same pattern of results. Yes, the relatively greater increases in earnings for female full-time workers show up broadly across all three earnings groups. Back in 1970, the relative earnings ratio of the highest earnings group was the lowest – at 53-54 percent. Over the next 45 years it went up by the greatest amount – by 25-28 percentage points. However, it still remains the most resistant to leading the rise. Among the lowest earnings group, the ratio rose by 20-23 points to a 2015 ratio of 84-85 percent, the highest ratio among the three groups. These changes are all highly statistically significant – see the last line in each table. But again, the full-time earnings gap has only about half closed.

4. Overview, Findings, and Conclusions

This paper calls for a rethink about how much of empirical income distribution analysis is performed. In an era when complex distributional changes have been occurring so that summary income inequality measures (such as, say, a Gini coefficient) cannot adequately capture what has been happening, more disaggregative measures of distributional change are much more useful. This paper offers a tool box of such disaggregative measures including income shares, population shares, quantile mean incomes and relative mean incomes. As can be seen from the illustrative empirical work, all of these tools can be usefully employed together to highlight different complementary aspects of distributional change and distributional differences, such as, say between men and women in the labour market. While the present study has illustrated such measures with the income (or earnings) distribution broken into three income groups (simply called Lower, Middle, and Higher), more extensive and refined breakdowns (e.g., Beach, 2018) can readily be applied given the development of the tools outlined in this paper.

Secondly, the paper calls for standard use of a statistical inferential framework for empirical distributional analysis, i.e., calculating standard errors that can be applied to all the disaggregative measures provided in the present tool box. This allows for, not just the study of the magnitudes of detailed distributional changes and differences, but also interpreting such calculated changes and differences in terms of their statistical significance (or prob-values if one wishes). This study has also developed a consistent inferential framework based on a quantile-function approach applied to each of the proposed tool box measures. In an era of broadly available public use microdata files such as from the Census and regular Labour Force Surveys, these standard error formulas can be readily calculated and implemented.

And thirdly, the paper argues for a broader use of median-based distributional statistics for detailed empirical work rather than the more readily available percentile-based statistics (such as quintiles and deciles) typically provided by official statistical agencies. Both sets of statistics can be readily grounded in a conventional statistical inference framework – as shown in the appendices of this study – but the median-based approach also allows for estimates of numbers of people (or population shares) in each income group and hence separate analysis of the degree of polarization within an income distribution and not just analysis of income differences and income inequality. It turns out, as well, that such an approach allows one to distinguish between a quantity (of workers) dimension and a price (income level or wage level) dimension. As shown in Beach (2016), this distinction can provide valuable insights into an interpretation of the underlying economic forces driving observed changes in income shares such as have occurred in the Canadian and U.S. labour markets over the last fifty years.

Several major empirical findings have been obtained when this tool box of measures has been applied to the distribution of workers' earnings in Canada over the 1970-2015 period. The empirical analysis has made use of Canadian Census public use microdata files for 1971, 1981, 1991, 2001, and 2006 (reporting earnings for the preceding years) and of monthly Labour Force Survey (LFS) public use microdata files for May of 2000, 2005, 2010, and 2015 (reporting earnings for a specified survey week). Several major results were found.

First, there has been a very substantial decline in the proportion of middle earners or middle-class workers in the Canadian labour market since the 1980s and a corresponding increase in the proportion of higher earners – for both male and female workers – and these changes are highly statistically significant. The result has been a growing degree of polarization in the labour market and a widening gap between top earnings levels and the rest of the workers in the labour market. The evidence is consistent with a declining fraction of middle-class workers in Canada.

Second, for male workers over the 1980s and 1990s, actual decreases in real earnings levels occurred over the lower and middle-class regions of the earnings distributions; these were quite substantial and historic and highly statistically significant. Since 2005, however, real earnings increases have occurred and been fairly broadly shared across all earnings groups and for both male and female workers.

Third, since the 1980s, a rising tide of a generally increasing real earnings level overall for the economy has not lifted all boats evenly. Relative-mean earnings ratios over the 1980-2005 period markedly declined for lower earnings and middle-class earnings workers, while very strongly rising for higher earnings workers. These changes are also highly statistically significant. The most consistent pattern of change since 1980 has been the quite marked decline in relative-mean earnings for middle-class workers, which has continued since 2005 and remains highly statistically significant.

Fourth, the increase in relative earnings of female compared to male full-time workers since 1970 has been broadly shared across all three earnings groups, but over a period of 45 years the gap has only been halved in size to about 20 percent.

The practical conclusions or recommendations that arise from these very strong and robust results are twofold. One, official statistical agencies (such as Statistics Canada or the U.S. Bureau of the Census) should provide distributional statistics on median-based measures of disaggregative change in addition to their current statistics based on quintile or decile breakdowns. And two, microdata users should use the formulas developed in this paper to compute standard errors for their various distributional statistics so they can base their analysis on more formal principles of statistical inference.

Table 1
Percentage of Male and Female Full-Time Workers by Earnings Level, Canada, 1970-2005
Census Data on Annual Earnings
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
1970	9.14 (0.175)	8.14 (0.313)
1980	9.35 (0.115)	8.51 (0.167)
1990	11.09 (0.130)	12.75 (0.103)
2000	13.33 (0.112)	13.24 (0.131)
2005	14.20 (0.113)	13.83 (0.127)
Change 1970-2005	+5.06 [24.3]	+5.68 [16.8]
Change 1980-2005	+4.85 [29.9]	+5.10 [24.3]
Middle-Class Earnings		
(within 50% of median)		
1970	74.34 (0.0523)	76.53 (0.0839)
1980	75.14 (0.0312)	75.21 (0.0467)
1990	71.46 (0.0128)	68.14 (0.0123)
2000	65.35 (0.0085)	65.52 (0.0088)
2005	62.78 (0.0108)	63.06 (0.0133)
Change 1970-2005	-11.56 [216.]	-13.47 [158.]
Change 1980-2005	-12.36 [374.]	-12.15 [250.]
Higher Earnings		
(above 200% of median)		
1970	6.73 (0.1452)	4.64 (0.2418)
1980	5.56 (0.0949)	4.60 (0.1326)
1990	6.46 (0.1013)	5.66 (0.0820)
2000	8.89 (0.0927)	7.83 (0.1047)
2005	10.14 (0.1010)	9.53 (0.1112)
Change 1970-2005	+3.41 [19.3]	+4.89 [18.4]
Change 1980-2005	+4.58 [33.0]	+4.93 [28.5]

Source: Statistics Canada, Census of Canada Individual PUMF files for 1971, 1981, 1991, 2001, and 2006.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

Table 2
Percentage of Male and Female Full-Time Workers by Earnings Level, Canada, 2000-2015
LFS Data on Weekly Earnings
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
2000	7.41 (0.137)	7.26 (0.151)
2005	8.12 (0.101)	8.06 (0.113)
2010	8.22 (0.077)	7.03 (0.079)
2015	7.89 (0.047)	6.55 (0.041)
Change 2000-2015	+0.48 [3.31]	-0.71 [4.53]
Change 2000-2010	+0.81 [5.15]	-0.23 [1.46]
Middle-Class Earnings		
(within 50% of median)		
2000	76.17 (0.066)	73.89 (0.092)
2005	75.06 (0.167)	73.05 (0.189)
2010	72.59 (0.244)	72.50 (0.268)
2015	73.41 (0.223)	71.41 (0.266)
Change 2000-2015	-2.76 [11.9]	-2.48 [8.81]
Change 2000-2010	-3.58 [14.16]	-1.39 [4.91]
Higher Earnings		
(above 200% of median)		
2000	4.79 (0.083)	6.16 (0.103)
2005	5.88 (0.135)	6.58 (0.159)
2010	5.91 (0.188)	6.73 (0.204)
2015	6.31 (0.178)	7.31 (0.207)
Change 2000-2015	+1.52 [7.75]	+1.15 [4.96]
Change 2000-2010	+1.12 [5.45]	+0.57 [2.49]

Source: Based on Statistics Canada, PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 3
Male and Female Earnings Shares of Full-Time Workers by Earnings Level, Canada, 1970-2005 (Percent)
Census Data on Annual Earnings
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
1970	2.52 (0.0780)	2.34 (0.1461)
1980	2.71 (0.0527)	2.40 (0.0775)
1990	2.96 (0.0598)	3.65 (0.0464)
2000	3.50 (0.0491)	3.59 (0.0590)
2005	3.40 (0.0444)	3.72 (0.0537)
Change 1970-2005	+0.88 [9.81]	+1.38 [8.86]
Change 1980-2005	+0.69 [10.01]	+1.32 [14.00]
Middle-Class Earnings		
(within 50% of median)		
1970	64.19 (0.226)	69.34 (0.410)
1980	67.62 (0.148)	68.44 (0.220)
1990	63.22 (0.137)	61.58 (0.109)
2000	55.43 (0.109)	57.45 (0.130)
2005	47.33 (0.102)	51.50 (0.124)
Change 1970-2005	-16.86 [68.0]	-17.84 [41.7]
Change 1980-2005	-20.19 [112.4]	-16.94 [67.1]
Higher Earnings		
(above 200% of median)		
1970	18.51 (0.259)	11.41 (0.451)
1980	14.19 (0.173)	10.59 (0.246)
1990	16.88 (0.187)	13.54 (0.148)
2000	22.51 (0.162)	18.14 (0.189)
2005	32.00 (0.158)	24.97 (0.188)
Change 1970-2005	+13.49 [44.4]	+13.56 [27.7]
Change 1980-2005	+17.81 [76.0]	+14.38 [46.4]

Source: Statistics Canada, Census of Canada Individual PUMF files for 1971, 1981, 1991, 2001, and 2006.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 4
Male and Female Earnings Shares of Full-Time Workers by Earnings Level, Canada, 2000-2015
LFS Data on Weekly Earnings
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
2000	2.80 (0.064)	2.82 (0.069)
2005	3.04 (0.046)	3.06 (0.051)
2010	3.09 (0.035)	2.67 (0.035)
2015	2.94 (0.021)	2.52 (0.018)
Change 2000-2015	+0.14 [2.09]	-0.30 [4.22]
Change 2000-2010	+0.29 [3.98]	-0.15 [1.94]
Middle-Class Earnings		
(within 50% of median)		
2000	68.18 (0.220)	64.52 (0.263)
2005	66.72 (0.323)	63.16 (0.360)
2010	63.16 (0.402)	61.53 (0.432)
2015	63.48 (0.346)	59.23 (0.389)
Change 2000-2015	-4.70 [11.5]	-5.29 [11.3]
Change 2000-2010	-5.02 [10.9]	-2.99 [5.91]
Higher Earnings		
(above 200% of median)		
2000	10.36 (0.153)	13.05 (0.188)
2005	13.16 (0.249)	14.59 (0.287)
2010	13.17 (0.341)	14.58 (0.367)
2015	14.37 (0.324)	15.77 (0.366)
Change 2000-2015	+3.51 [9.80]	+2.72 [6.60]
Change 2000-2010	+2.31 [6.18]	+1.53 [3.93]

Source: Based on Statistics Canada, PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 5
Male and Female Earnings Shares of Full-Time Workers by Earnings Level, Canada, 1970-2005 (Percent)
Census Data on Annual Earnings
(Percentile-Based Estimates)

	Males	Females
Lower Earnings		
(bottom 20%)		
1970	8.32 (0.145)	9.05 (0.279)
1980	8.62 (0.096)	8.88 (0.149)
1990	7.60 (0.068)	7.48 (0.084)
2000	6.80 (0.062)	7.01 (0.076)
2005	5.91 (0.053)	6.60 (0.065)
Change 1970-2005	-2.41 [15.7]	-2.45 [8.54]
Change 1980-2005	-2.71 [24.7]	-2.28 [14.0]
Middle-Class Earnings		
(middle 60%)		
1970	53.93 (0.294)	56.33 (0.566)
1980	55.77 (0.196)	56.86 (0.304)
1990	55.20 (0.148)	56.53 (0.192)
2000	53.89 (0.149)	55.74 (0.187)
2005	48.48 (0.139)	52.68 (0.171)
Change 1970-2005	-5.45 [16.8]	-3.65 [6.19]
Change 1980-2005	-7.29 [30.3]	-4.18 [12.0]
Higher Earnings		
(top 20%)		
1970	37.75 (0.295)	34.62 (0.567)
1980	35.61 (0.196)	34.26 (0.305)
1990	37.20 (0.149)	35.98 (0.194)
2000	39.32 (0.152)	37.26 (0.190)
2005	45.62 (0.142)	40.72 (0.175)
Change 1970-2005	+7.87 [24.1]	+6.10 [10.3]
Change 1980-2005	+10.01 [41.4]	+6.46 [18.4]

Source: Statistics Canada, Census of Canada Individual PUMF files for 1971, 1981, 1991, 2001, and 2006.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix C.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 6
Male and Female Earnings Shares of Full-Time Workers by Earnings Level, Canada, 2000-2015
LFS Data on Weekly Earnings
(Percentile-Based Estimates)

	Males	**Females**
Lower Earnings		
<u>(bottom 20%)</u>		
2000	9.83 (0.180)	9.67 (0.199)
2005	9.55 (0.179)	9.35 (0.192)
2010	9.49 (0.175)	9.76 (0.182)
2015	9.78 (0.124)	10.23 (0.134)
Change 2000-2015	-0.05 [0.19]	+0.56 [2.33]
Change 2000-2010	-0.34 [1.35]	+0.09 [0.33]
Middle-Class Earnings		
<u>(middle 60%)</u>		
2000	56.43 (0.371)	56.11 (0.437)
2005	55.94 (0.380)	55.42 (0.424)
2010	55.68 (0.383)	55.09 (0.409)
2015	54.91 (0.277)	54.29 (0.303)
Change 2000-2015	-1.52 [3.28]	-1.82 [3.44]
Change 2000-2010	-0.75 [1.41]	-1.02 [1.70]
Higher Earnings		
<u>(top 20%)</u>		
2000	33.74 (0.373)	34.22 (0.442)
2005	34.51 (0.383)	35.23 (0.429)
2010	34.82 (0.387)	35.15 (0.414)
2015	35.30 (0.281)	35.49 (0.307)
Change 2000-2015	+1.56 [3.35]	+1.27 [2.36]
Change 2000-2010	+1.08 [2.01]	+0.93 [1.54]

Source: Based on Statistics Canada, PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix C.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

Table 7
Conditional Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 1970-2005
Census Data on Annual Earnings (real 2015$)
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
1970	15,316. (181.0)	9476. (227.0)
1980	18,909. (133.9)	11,863. (150.4)
1990	17,245. (146.5)	12,618. (58.9)
2000	17,163. (96.4)	12,820. (84.3)
2005	17,832. (90.5)	14,007. (73.5)
Change 1970-2005	+2516. [12.43]	+4540. [19.03]
Change 1980-2005	-1077. [6.67]	+2144. [12.80]
Middle-Class Earnings		
(within 50% of median)		
1970	47,967. (135.1)	29,838. (143.6)
1980	58,709. (104.1)	38,279. (99.3)
1990	57,162. (113.9)	38,833. (63.7)
2000	55,444. (102.2)	41,456. (88.2)
2005	56,146. (110.8)	42,530. (93.6)
Change 1970-2005	+8179. [46.81]	+12,962. [74.04]
Change 1980-2005	-2563. [16.86]	+4251. [31.15]
Higher Earnings		
(above 200% of median)		
1970	152,789. (1159.)	80,982. (1017.)
1980	166,498. (810.6)	96,841. (539.2)
1990	168,831. (781.1)	105,440. (372.1)
2000	165,511. (531.4)	109,533. (322.5)
2005	235,026. (1179.)	136,447. (564.7)
Change 1970-2005	+82,237. [49.75]	+55,465. [47.68]
Change 1980-2005	+68,528. [47.90]	+39,606. [50.73]

Source: Statistics Canada, Census of Canada Individual PUMF files for 1971, 1981, 1991, 2001, and 2006.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

Table 8
Conditional Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 2000-2015
LFS Data on Weekly Earnings (real 2015$)
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
2000	418.1 (1.75)	328.8 (1.16)
2005	415.8 (1.18)	334.3 (0.91)
2010	445.5 (0.87)	364.6 (0.75)
2015	452.9 (0.59)	384.0 (0.36)
Change 2000-2015	+34.8 [18.83]	+55.2 [45.44]
Change 2000-2010	+27.4 [14.00]	+35.8 [25.92]
Middle-Class Earnings		
(within 50% of median)		
2000	990.4 (2.33)	739.2 (2.09)
2005	987.1 (2.58)	761.4 (2.36)
2010	1031.1 (3.10)	814.7 (2.71)
2015	1051.1 (2.54)	827.8 (2.35)
Change 2000-2015	+60.7 [17.59]	+88.6 [28.16]
Change 2000-2010	+40.7 [10.50]	+75.5 [22.06]
Higher Earnings		
(above 200% of median)		
2000	2508.7 (7.89)	1793.3 (4.29)
2005	2485.4 (10.17)	1952.6 (8.59)
2010	2640.9 (15.56)	2079.5 (10.70)
2015	2768.1 (15.73)	2153.0 (11.04)
Change 2000-2015	+259.4 [14.74]	+359.7 [30.36]
Change 2000-2010	+132.2 [7.58]	+286.2 [24.82]

Source: Based on Statistics Canada, PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

Table 9
Conditional Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 1970-2005
Census Data on Annual Earnings (real 2015$)
(Percentile-Based Estimates)

	Males	**Females**
Lower Earnings		
(bottom 20%)		
1970	23,122. (402.7)	14,901. (459.7)
1980	28,130. (313.5)	18,677. (313.9)
1990	24,544. (219.3)	16,492. (185.3)
2000	22,217. (201.0)	16,570. (180.1)
2005	21,992. (196.6)	17,196. (170.2)
Change 1970-2005	-1129. [2.80]	+2294. [4.99]
Change 1980-2005	-6138. [16.59]	+1481. [4.15]
Middle-Class Earnings		
(middle 60%)		
1970	49,931. (272.4)	30,919. (310.4)
1980	60,634. (213.0)	39,867. (213.2)
1990	59,445. (159.0)	41,529. (140.7)
2000	58,706. (162.4)	43,918. (147.1)
2005	60,170. (172.0)	45,719. (148.5)
Change 1970-2005	+10,239. [31.76]	+14,800. [42.90]
Change 1980-2005	-464. [1.69]	+5852. [22.53]
Higher Earnings		
(top 20%)		
1970	104,847. (818.8)	56,988. (932.9)
1980	116,151. (640.8)	72,041. (641.3)
1990	120,179. (481.9)	79,300. (428.4)
2000	128,490. (496.9)	88,072. (450.3)
2005	169,863 (528.8)	106,025. (456.3)
Change 1970-2005	+65,016. [66.74]	+49,037. [47.23]
Change 1980-2005	+53,712. [64.65]	+33,984. [43.18]

Source: Statistics Canada, Census of Canada Individual PUMF files for 1971, 1981, 1991, 2001, and 2006.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix C.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 10
Conditional Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 2000-2015
LFS Data on Weekly Earnings (real 2015$)
(Percentile-Based Estimates)

	Males	**Females**
Lower Earnings		
(bottom 20%)		
2000	540.3 (9.95)	408.7 (8.40)
2005	531.3 (9.94)	412.3 (8.47)
2010	559.8 (10.36)	457.1 (8.75)
2015	572.6 (7.51)	482.7 (6.69)
Change 2000-2015	+32.4 [2.59]	+74.0 [6.89]
Change 2000-2010	+19.5 [1.36]	+48.4 [3.99]
Middle-Class Earnings		
(middle 60%)		
2000	1044.9 (6.84)	794.0 (6.17)
2005	1037.6 (7.03)	814.6 (6.22)
2010	1104.3 (7.57)	890.9 (6.54)
2015	1129.6 (5.61)	922.9 (5.03)
Change 2000-2015	+84.7 [9.57]	+128.9 [16.19]
Change 2000-2010	+59.4 [5.82]	+96.9 [10.78]
Higher Earnings		
(top 20%)		
2000	1874.4 (20.61)	1451.8 (18.72)
2005	1919.9 (21.26)	1557.0 (18.87)
2010	2068.9 (22.95)	1690.7 (19.87)
2015	2152.2 (17.05)	1774.5 (15.31)
Change 2000-2015	+277.8 [10.38]	+322.7 [13.34]
Change 2000-2010	+194.5 [6.31]	+238.9 [8.75]

Source: Based on Statistics Canada, PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix C.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

Table 11
Relative Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 1970-2005
Census Data on Annual Earnings
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
1970	.2757 (.00326)	.2875 (.00639)
1980	.2898 (.00205)	.2820 (.00358)
1990	.2669 (.00227)	.2863 (.00134)
2000	.2626 (.00148)	.2711 (.00178)
2005	.2394 (.00122)	.2690 (.00141)
Change 1970-2005	-.0363 [10.43]	-.0185 [2.63]
Change 1980-2005	-.0504 [21.14]	-.0130 [3.38]
Middle-Class Earnings		
(within 50% of median)		
1970	.8635 (.00248)	.9060 (.00436)
1980	.8999 (.00160)	.9100 (.00236)
1990	.8847 (.00176)	.9037 (.00144)
2000	.8482 (.00156)	.8768 (.00187)
2005	.7539 (.00149)	.8167 (.00180)
Change 1970-2005	-.1096 [38.43]	-.0893 [18.95]
Change 1980-2005	-.1460 [66.92]	-.0933 [31.44]
Higher Earnings		
(above 200% of median)		
1970	2.7504 (.02086)	2.4591 (.03088)
1980	2.5522 (.01242)	2.3022 (.01282)
1990	2.6130 (.01209)	2.3922 (.00844)
2000	2.5321 (.00813)	2.3167 (.00682)
2005	3.1588 (.01583)	2.6201 (.01084)
Change 1970-2005	+.4054 [15.48]	+.1611 [4.92]
Change 1980-2005	+.6036 [30.00]	+.3179 [18.93]

Source: Statistics Canada, Census of Canada Individual PUMF files for 1971, 1981, 1991, 2001, and 2006.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 12
Relative Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 2000-2015
LFS Data on Weekly Earnings
(Median-Based Estimates)

	Males	**Females**
Lower Earnings		
(below 50% of median)		
2000	.3779 (.00158)	.3884 (.00137)
2005	.3744 (.00106)	.3797 (.00103)
2010	.3759 (.00073)	.3798 (.00078)
2015	.3726 (.00048)	.3847 (.00036)
Change 2000-2015	-.0053 [3.17]	-.0037 [2.61]
Change 2000-2010	-.0020 [1.15]	-.0086 [5.45]
Middle-Class Earnings		
(within 50% of median)		
2000	.8951 (.00210)	.8732 (.00247)
2005	.8889 (.00233)	.8646 (.00268)
2010	.8701 (.00262)	.8487 (.00282)
2015	.8647 (.00209)	.8294 (.00236)
Change 2000-2015	-.0304 [10.24]	-.0438 [12.81]
Change 2000-2010	-.0250 [7.45]	-.0245 [6.53]
Higher Earnings		
(above 200% of median)		
2000	2.2672 (.00713)	2.1185 (.00506)
2005	2.2381 (.00916)	2.2173 (.00976)
2010	2.2284 (.01313)	2.1664 (.01115)
2015	2.2773 (.01294)	2.1573 (.01107)
Change 2000-2015	+.0101 [0.68]	+.0388 [3.19]
Change 2000-2010	-.0388 [2.60]	+.0479 [3.91]

Source: Based on Statistics Canada, PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix B.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 13
Relative Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 1970-2005
Census Data on Annual Earnings
(Percentile-Based Estimates)

	Males	**Females**
Lower Earnings		
(bottom 20%)		
1970	.4162 (.00725)	.4526 (.01396)
1980	.4312 (.00481)	.4440 (.00746)
1990	.3798 (.00339)	.3742 (.00420)
2000	.3399 (.00307)	.3505 (.00381)
2005	.2953 (.00264)	.3302 (.00327)
Change 1970-2005	-.1209 [15.67]	-.1224 [8.54]
Change 1980-2005	-.1359 [24.77]	-.1138 [13.97]
Middle-Class Earnings		
(middle 60%)		
1970	.8988 (.00480)	.9388 (.00943)
1980	.9294 (.00327)	.9478 (.00507)
1990	.9200 (.00246)	.9422 (.00319)
2000	.8981 (.00248)	.9289 (.00311)
2005	.8079 (.00231)	.8779 (.00285)
Change 1970-2005	-.0909 [16.77]	-.0609 [6.19]
Change 1980-2005	-.1215 [30.34]	-.0699 [12.02]
Higher Earnings		
(top 20%)		
1970	1.8873 (.01474)	1.7309 (.02833)
1980	1.7805 (.00982)	1.7127 (.01525)
1990	1.8601 (.00746)	1.7992 (.00972)
2000	1.9658 (.00760)	1.8628 (.00952)
2005	2.2809 (.00710)	2.0360 (.00876)
Change 1970-2005	+.3936 [24.05]	+.3051 [10.29]
Change 1980-2005	+.5004 [41.29]	+.3233 [18.38]

Source: Statistics Canada, Census of Canada Individual PUMF files for 1971, 1981, 1991, 2001, and 2006.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix C.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

Table 14
Relative Mean Earnings of Full-Time Male and Female Workers by Earnings Level, Canada, 2000-2015
LFS Data on Weekly Earnings
(Percentile-Based Estimates)

	Males	**Females**
Lower Earnings		
(bottom 20%)		
2000	.4913 (.00900)	.4833 (.00993)
2005	.4777 (.00895)	.4674 (.00962)
2010	.4747 (.00874)	.4881 (.00912)
2015	.4891 (.00618)	.5112 (.00671)
Change 2000-2015	-.0022 [0.19]	+.0279 [2.33]
Change 2000-2010	-.0166 [1.32]	+.0048 [0.36]
Middle-Class Earnings		
(middle 60%)		
2000	.9406 (.00618)	.9353 (.00728)
2005	.9323 (.00633)	.9236 (.00706)
2010	.9280 (.00638)	.9181 (.00681)
2015	.9152 (.00461)	.9048 (.00504)
Change 2000-2015	-.0254 [3.29]	-.0305 [3.44]
Change 2000-2010	-.0126 [1.42]	-.0172 [1.73]
Higher Earnings		
(top 20%)		
2000	1.6871 (.01863)	1.7109 (.02211)
2005	1.7253 (.01914)	1.7617 (.02143)
2010	1.7412 (.01937)	1.7576 (.02070)
2015	1.7652 (.01403)	1.7745 (.01534)
Change 2000-2015	+.0781 [3.35]	+.0636 [2.36]
Change 2000-2010	+.0541 [2.01]	+.0467 [1.54]

Source: Based on Statistics Canada, PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors as specified in Appendix C.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 15
Female-to-Male Earnings Ratios for Full-Time Workers Based on Mean and Median Earnings, Canada, 1970-2015

	Relative Mean Earnings	Relative Median Earnings
1970	.5928	.6203
1980	.6448	.6567
1990	.6822	.6944
2000	.7233	.7459
2005	.6993	.7551
2000	.7650	.7500
2005	.7930	.7813
2010	.8100	.8024
2015	.8211	.7977

Source: See Appendix Tables D1 – D4.

Note: Top panel of figures from Census data; bottom panel of figures from LFS data.

Table 16
Female-to-Male Earnings Ratios for Full-Time Workers by Earnings Level, Canada, 1970-2015
(Median-Based Estimates)

	Lower Earners	**Middle Earners**	**Higher Earners**
1970	.6187 (.0165)	.6221 (.00347)	.5300 (.00778)
1980	.6274 (.00911)	.6520 (.00205)	.5816 (.00430)
1990	.7317 (.00709)	.6968 (.00178)	.6245 (.00363)
2000	.7470 (.00646)	.7477 (.00210)	.6618 (.00288)
2005	.7855 (.00573)	.7575 (.00224)	.5806 (.00378)
2000	.7864 (.00431)	.7464 (.00275)	.7148 (.00282)
2005	.8040 (.00316)	.7714 (.00312)	.7856 (.00472)
2010	.8184 (.00232)	.7901 (.00354)	.7874 (.00616)
2015	.8479 (.00136)	.7876 (.00294)	.7778 (.00595)
Change 1970-2015	+.2279 [13.82]	+.1655 [36.42]	+.2478 [25.30]

Source: See Tables 7 and 8.

Note: Top panel of figures from Census data; bottom panel of figures from LFS data.

Table 17
Female-to-Male Earnings Ratios for Full-Time Workers by Earnings Level, Canada, 1970-2015
(Percentile-Based Estimates)

	Lower Earners	**Middle Earners**	**Higher Earners**
1970	.6445 (.0228)	.6192 (.00708)	.5435 (.00986)
1980	.6640 (.0134)	.6575 (.00421)	.6202 (.00650)
1990	.6719 (.0096)	.6986 (.00302)	.6598 (.00444)
2000	.7458 (.0106)	.7481 (.00325)	.6854 (.00439)
2005	.7819 (.0104)	.7598 (.00329)	.6242 (.00332)
2000	.7564 (.0209)	.7599 (.00772)	.7745 (.0133)
2005	.7760 (.0216)	.7851 (.00801)	.8110 (.0133)
2010	.8165 (.0217)	.8068 (.00810)	.8172 (.0132)
2015	.8430 (.0161)	.8170 (.00602)	.8245 (.00966)
Change 1970-2015	+.1985 [7.11]	+.1978 [21.29]	+.2810 [20.36]

Source: See Tables 9 and 10.

Note: Top panel of figures from Census data; bottom panel of figures from LFS data.

References

Acemoglu, D.H., D. Autor, G.H. Hanson, and B. Price (2016) "Import Competition and the Great U.S. Employment Sag of the 2000s", *Journal of Labor Economics* 34, S141-S198.

Aitchison, J., and J.A.C. Brown (1957) *The Lognormal Distribution with Special Reference to Its Uses in Economics*. Cambridge: University of Cambridge Department of Applied Economics Monograph No. 5; Cambridge University Press.

Autor, D.H., D. Dorn, and G.H. Hanson (2013) "The Geography of Trade and Technology Shocks in the United States", *American Economic Review* 103, 220-225.

Beach, C.M. (2016) "Changing Income Inequality: A Distributional Paradigm for Canada", *Canadian Journal of Economics* 49, 1229-1292.

__________ (2017) "Have Middle-Class Earnings Risen in Canada? A Statistical Inference Approach", Queen's University, Department of Economics Discussion Paper No. 1393, November.

__________ (2018) "Distributional Gains of Near Higher Earners", Queen's University, Department of Economics Discussion Paper No. 1398, February.

Beach, C.M., K.V. Chow, J.P. Formby, and G.A. Slotsve (1994) "Statistical Inference for Decile Means", *Economics Letters* 45, 161-167.

Beach, C.M. and R. Davidson (1983), "Distribution-Free Statistical Inference with Lorenz Curves and Income Shares", *Review of Economic Studies* 50, 723-735.

Beaudry, P., and D.A. Green (2006) "Changes in U.S. Wages 1976-2000: Ongoing Skill Bias or Major Technological Change", *Journal of Labor Economics* 23, 609-648.

Bergin, J. (2015) *Mathematics for Economists with Applications*. London and New York: Routledge.

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.J. Slottje, Greenwich, CT: JAI Press, 49-82.

__________ (1991) "Changes in the U.S. Earnings Distribution in the 1980's", *Applied Economics* 23(3), 425-433.

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

Bouderbat, B., T. Lemieux, and W.C. Riddell (2010) "The Evolution of Returns to Human Capital in Canada, 1980-2005", *Canadian Public Policy* 36, 63-89.

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

Di Nardo, J., N.M. Fortin, and T. Lemieux (1996) “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach”, *Econometrica* 64, 1001-1044.

Goos, M., A. Manning, and A. Salomons (2014) “Explaining Job Polorization: Routine-Biased Technological Change and Offshoring”, *American Economic Review* 104, 2509-2526.

Green, D.A., W.C. Riddell, and F. St-Hilaire (2016) “Income Inequality in Canada: Driving Forces, Outcomes and Policy”, in *Income Inequality: The Canadian Story*, ed. by D.A. Green, W.C. Riddell, and F. St-Hilaire. Montreal: The Institute for Research on Public Policy.

Green, D.A., and B.M. Sand (2015) “Has the Canadian Labour Market Polarized?”, *Canadian Journal of Economics* 48, 612-645.

Jenkins, S. (1991) “The Measurement of Income Inequality”, in *Economic Inequality and Poverty: International Perspectives*, ed. By Lars Osberg, Armonk, New York: M.E. Sharpe, Inc.

Katz, L.F., and K.M. Murphy (1992) “Changes in Relative Wages, 1963-1987: Supply and Demand Factors”, *Quarterly Journal of Economics* 107, 35-78.

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.

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.

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

Technical Appendix A

Percentile-Based Statistics

This appendix sets out what may be viewed as the conventional approach to implementing statistical inference to disaggregative distributional measures such as income shares or quantile mean income levels, arising from the work of Beach and Davidson (1983) and Beach et al. (1994). This approach seems not to have been actually used much in empirical distributional studies, perhaps in part because the formulas in the former paper may not seem very user-friendly to regular empirical practitioners. This appendix seeks to make this approach more accessible to such users. But it also sets out a foil or contrast to an alternative "quantile function" approach to disaggregative statistical inference provided in the following two appendices.

A.1 Standard Errors for Quantile Means

The exposition of how (asymptotic) standard errors of quantile income shares are obtained follows two steps.

Step 1: Asymptotic Distribution of Cumulative Quantile Means

A formal derivation of the key result in Step 1 is provided in Beach and Davidson (1983), but some exposition and notation will be needed here. Suppose there are K (ordered) percentile income groups indexed by i, from the lowest-income group $i=1$ up to the top group K. In the case of decile income groups, $K=10$; for quintiles, $K=5$. Let p_i represent the (cumulative) proportion of the ordered income groups. So for deciles, $p_1 = 0.1, p_2 = 0.2, \ldots, p_9 = 0.9$. The income cut-offs that divide the different income groups will be called $\xi_1, \xi_2, \ldots, \xi_{K-1}$, and the *cumulative* quantile means are $\gamma_1, \gamma_2, \dots, \gamma_K$ where $\gamma_i = E(Y \mid Y \leq \xi_i)$ for incomes less than or equal to ξ_i and $\gamma_K = \mu$ the overall mean of the distribution of sample incomes Y. Cumulative quantile variances are denoted $\lambda_i^2 = var(Y \mid Y \leq \xi_i)$ and the overall variance of the distribution in $\lambda_K^2 = \sigma^2$. Let the (K)-element vector of sample estimates of the (rescaled) cumulative means be

$$\hat{G} = (p_1\hat{\gamma}_1, \dots, p_K\hat{\gamma}_K)' .$$

So $\hat{G} = P\hat{\gamma}$ for vector $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_K)'$ and P = Diag $(p_1, \dots, p_K)$. Then Beach and Davidson (1983, Theorem 1) establish that, under general conditions, $\hat{G}$ is asymptotically joint normally distributed with mean $G = (p_1\gamma_1, \dots, p_K\gamma_K)'$ and variance-covariance matrix

$\Omega = [\omega_{ij}]$ where

$$\omega_{ij} = p_i\big[\lambda_i^2 + (1 - p_i)(\xi_i - \gamma_i)\big(\xi_j - \gamma_j\big) + (\xi_i - \gamma_i)(\gamma_j - \gamma_i)\big]$$

$$\text{for } i \leq j = 1, \dots, K. \qquad \text{(a1)}$$

When $i = j$, the corresponding variance of $p_i\hat{\gamma}_i$ is

$$\omega_{ii} = p_i[\lambda_i^2 + (1 - p_i)(\xi_i - \gamma_i)^2] . \qquad \text{(a2)}$$

Step 2: Asymptotic Distribution of Quantile Means

Now consider the K-vector of estimated *quantile* mean incomes

$$\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_K)' .$$

where μ_i is the mean of the i'th income group. It can be seen that the vector $\hat{\mu}$ is a linear function of $\hat{\gamma}$ and hence of $\hat{G}$. Beach et al. (1994) then show that vector $\hat{\mu}$ is also jointly asymptotically normally distributed with mean $\mu = (\mu_1, \dots, \mu_K)'$ and (asymptotic) variance-covariance matrix

$$V = R\,\Omega\,R'$$

where $R = (PA)^{-1}$ and A is such that

$$\hat{\gamma} = A\hat{\mu} \text{ or conversely } \hat{\mu} = A^{-1}\hat{\gamma} ,$$

so that

$$R = P^{-1}A^{-1}$$

$$= \begin{bmatrix} \frac{1}{p_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{p_K} \end{bmatrix} \bullet \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & 0 & \cdots & 0 & 0 \\ 0 & -2 & 3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & & -(K-1) & K \end{bmatrix} .$$

Since A^{-1} and hence *R* are lower-triangular, one can make use of a simple scaler algorithm for the (asymptotic) variances and standard errors of the sample quantile means. In the case of a decile breakdown of income groups (*K* = *10*),

$$Asy.var(\hat{\mu}_1) = 100\ \omega_{11} \tag{a3}$$

and $$Asy.var(\hat{\mu}_i) = 100\left(\omega_{ii} + \omega_{j-i,i-1} - 2\omega_{i,i-1}\right), \qquad i = 2, \dots, 10 . \tag{a4}$$

Alternatively, in terms of cumulative means,

$$\hat{\mu}_1 = \hat{\gamma}_1$$

and $$\hat{\mu}_i = \frac{p_i\hat{\gamma}_i - p_{i-1}\hat{\gamma}_{i-1}}{p_i - p_{i-1}} = i\hat{\gamma}_i - (i-1)\,\hat{\gamma}_{i-1}\,, i = 2, \dots, 10\,,$$

so that $$Asy.var(\hat{\mu}_1) = Asy.var(\hat{\gamma}_1) \tag{a5}$$

and $$Asy.var(\hat{\mu}_i) = i^2 \bullet Asy.var(\hat{\gamma}_i) + (i-1)^2 \bullet Asy.var(\hat{\gamma}_{i-1})$$

$$-2i(i-1) \bullet Asy.cov(\hat{\gamma}_i, \hat{\gamma}_{i-1}), \qquad i = 2, \dots, 10 \tag{a6}$$

where $$Asy.var(\hat{\gamma}_i) = \omega_{ii} \,/\, p_i^2 \tag{a7}$$

and $$Asy.cov(\hat{\gamma}_i, \hat{\gamma}_{i-1}) = \omega_{i,i-1} \,/\, p_i p_{i-1}\,. \tag{a8}$$

Note, incidentally, that all these results are distribution-free in that the (asymptotic) variances and covariances of both the cumulative and quantile sample means depend only on the first and second moments of the underlying sampling distribution, and these can be estimated consistently without having to know the underlying distribution itself.

A.2 Standard Errors for Income Shares

Now the (cumulative) income shares or Lorenz curve ordinates for income group *i* are given by

$$\widehat{LC}_i = p_i\hat{\gamma}_i \,/\, \hat{\mu} \qquad i = 1, \ldots, K\,.$$

Then Beach and Davidson (1983, Theorem 2) also establish that the set of Lorenz curve ordinates are also asymptotically joint normally distributed with (asymptotic) variance-covariance matrix $\theta = [\theta_{ij}]$ where

$$\theta_{ij} = \left(\frac{1}{\mu^2}\right)\omega_{ij} + \left(\frac{p_i\gamma_i}{\mu^2}\right)\left(\frac{p_j\gamma_j}{\mu^2}\right)\sigma^2 - \left(\frac{p_i\gamma_i}{\mu^3}\right)\omega_{j,K} - \left(\frac{p_j\gamma_j}{\mu^3}\right)\omega_{i,K}$$

$$\text{for } i \le j = 1, \ldots, K-1\,. \tag{a9}$$

In the case of (asymptotic) variances where *i* = *j*,

$$\theta_{ii} = \left(\frac{p_i}{\mu^2}\right)[\lambda_i^2 + (1 - p_i)(\xi_i - \gamma_i)^2] + \left(\frac{p_i\gamma_i}{\mu^2}\right)^2\sigma^2$$

$$-\,2\left(\frac{p_i^2\gamma_i}{\mu^3}\right)[\lambda_i^2 + (\mu - \gamma_i)(\xi_i - \gamma_i)]\,. \tag{a10}$$

Now the quantile income share IS_i is simply the vertical difference between its Lorenz curve ordinate and its previous Lorenz curve ordinate

$$IS_i = LC_i - LC_{i-1} \qquad i = 1, \ldots, K$$

where $LC_0 = 0$ and $LC_K = 1.0$. Hence,

$$\widehat{IS}_i = \widehat{LC}_i - \widehat{LC}_{i-1}$$

so that

$$Asy.var(\widehat{IS}_i) = Asy.var(\widehat{LC}_i) + Asy.var(\widehat{LC}_{i-1}) - 2Asy.cov(\widehat{LC}_i, \widehat{LC}_{i-1}) \tag{a11}$$

$$= \theta_{ii} + \theta_{i-1,i-1} - 2\theta_{i,i-1} \qquad \text{for } i = 2, \ldots, K\,.$$

and $Asy.var(\widehat{IS}_1) = \theta_{11}$.

Thus the estimated (asymptotic) variance of $\widehat{IS}_i$ is

$$Asy\hat{.}var(\widehat{IS}_i) = \hat{\theta}_{ii} + \hat{\theta}_{i-1,i-1} - 2\hat{\theta}_{i,i-1} \qquad \text{for } i = 2, \ldots, K \qquad \text{(a12)}$$

and $Asy\hat{.}var(\widehat{IS}_1) = \hat{\theta}_{11}$

where all the unknowns in eq. (a9) and (a10) are replaced by their sample estimates. The standard error, then, of $\widehat{IS}_i$ is

$$S.E.(\widehat{IS}_i) = \left[\frac{Asy\hat{.}var(\widehat{IS}_i)}{N}\right]^{1/2} . \qquad \text{(a13)}$$

Obviously, since the population shares are known by construction and not estimated, they have no standard errors.

A.3 Standard Errors for Relative Quantile Means

Once the formula for the standard error of income shares has been established, deriving that of relative-mean-income ratios ($\hat{\mu}_i$ / $\hat{\mu}$) is straightforward. Since

$$IS_i = (p_i - p_{i-1})\mu_i / \mu ,$$

then

$$\frac{\mu_i}{\mu} = \left(\frac{1}{p_i - p_{i-1}}\right) IS_i = \mathrm{K} \bullet IS_i \qquad \text{(a14)}$$

for K income shares. So

$$\frac{\hat{\mu}_i}{\hat{\mu}} = K \bullet I\hat{S}_i .$$

Hence, $S.E.(\hat{\mu}_i / \hat{\mu}) = K \bullet S.E.(I\hat{S}_i)$. (a15)

A.4 Standard Errors of Quantile Mean Differences

By quantile mean differences is meant the income gap $\mu_i - \mu_j$ between the i'th and j'th quantile mean incomes, for $i \neq j$. Again, the derivation of the standard error of $\hat{\mu}_i - \hat{\mu}_j$ depends on the development above in section A.1. More specifically, since

$$Asy.var(\hat{\mu}) = \mathrm{V} = \mathrm{R}\,\Omega\,\mathrm{R}' \tag{a16}$$

for the full vector $\hat{\mu}$ of sample quantile means, then the (asymptotic) variance-covariance matrix of $\hat{\mu}_i$ and $\hat{\mu}_j$ is

$$\begin{bmatrix} v_{ii} & v_{ij} \\ v_{ji} & v_{jj} \end{bmatrix} \quad \text{for } i \neq j \text{ ,}$$

so that

$$Asy.var(\hat{\mu}_i - \hat{\mu}_j) = v_{ii} + v_{jj} - 2v_{ij} \quad \text{since } v_{ji} = v_{ij} \text{ .} \tag{a17}$$

Consequently,

$$S.E.(\hat{\mu}_i - \hat{\mu}_j) = \left[As\hat{y.}var(\hat{\mu}_i - \hat{\mu}_j) \,/\, N\right]^{1/2}$$

$$= \left[\hat{v}_{ii} + \hat{v}_{jj} - 2\hat{v}_{ij}\right]^{1/2} / \sqrt{N}, \tag{a18}$$

where once again the unknown variances and covariances are estimated by their sample values.

Technical Appendix B

Median-Based Statistics Based on the Quantile Function Approach

B.1 Quantile Function Approach to Calculating Standard Errors

B.1.1 Deriving the Formulas

In this approach, the different income groups are separated by cut-off income levels that are expressed, not in percentile terms, but in terms of some fraction or multiple of the (sample) median income level. For illustrative purposes, consider three such income groups – lower incomes (L), middle incomes (M), and higher incomes (H). Middle class incomes, for example, are often expressed as those incomes between 50 percent and 150 percent of the median. If the sample median is $\hat{\xi}$, then the cut-offs are at $0.5\hat{\xi}$ and $1.5\hat{\xi}$ with

lower incomes less than $0.5\hat{\xi}$

middle incomes between $0.5\hat{\xi}$ and $1.5\hat{\xi}$,

and higher incomes greater than $1.5\hat{\xi}$.

The three income groups do not need to be exhaustive (i.e., cover the entire range of income levels in the distribution). For example, the top income group could run from $c\hat{\xi}$ and above, where $c > b$. Indeed, in the illustrative empirical work in Section 3 of this paper, this is the case where $c = 2.0$ or twice the median.

If *f(•)* is a specified (continuously differentiable) income distribution density function, then the share of recipients, or population share (*PS*), is given by

$$PS = \int_R f(y)dy \qquad \text{(b1)}$$

and the share of incomes, or income share (*IS*), is given by

$$IS = \int_R \frac{1}{\mu} yf(y)dy \qquad \text{where } \mu \equiv E(y) \tag{b2}$$

is the mean income of the distribution and *R* is the range of income levels defining each income group. In contrast to the percentile-based approach where each of the population shares is given as, say, 0.10 and 0.20 and hence is not estimated, now the integral bounds are functions of $\hat{\xi}$, the sample median, and hence *both* the income share and the population share are estimated and thus random variables.

Consider first an illustrative case of the middle-income group population share. In this case, the integral bounds are $0.5\hat{\xi}$ and $1.5\hat{\xi}$. So PS_M is estimated as the proportion of sample observations between $0.5\hat{\xi}$ and $1.5\hat{\xi}$ in the ordered sample of incomes.

Thus $$P\hat{S}_M = \int_{0.5\hat{\xi}}^{1.5\hat{\xi}} f(y)dy \ . \tag{b3}$$

The approach taken to establish the (asymptotic) distribution of the random variable $P\hat{S}_M$ is based on recognizing that this is a function of the sample median whose (asymptotic) distribution is well known. More specifically, under fairly broad conditions $\sqrt{N}\,(\hat{\xi} - \xi)$ has a limiting normal distribution with mean zero and variance

$$Asy.var(\hat{\xi}) = (.5)(.5) \,/\, [f(\xi)]^2 \equiv \theta(\xi)^2 \tag{b4}$$

where *N* is the sample size (Rao, 1965, p. 423). Hence, the (asymptotic) standard error of $\hat{\xi}$ is

$$Asy.S.E.(\hat{\xi}) = \left[As\hat{y.}var(\hat{\xi}) \,/\, N\right]^{1/2}$$

$$= (.5)/\left[f(\hat{\xi}) \bullet \sqrt{N}\right]. \tag{b5}$$

To link the share formulas to the median, recall from Rao (1965, p. 385) that, if $\hat{\xi}$ has a limiting normal distribution with (asymptotic) variance given by (b4) and if $g(\hat{\xi})$ is a continuous differentiable function of $\hat{\xi}$ with a first derivative $g'(\hat{\xi}) \equiv dg(\hat{\xi}) \,/\, d\hat{\xi}$, then the statistic $g(\hat{\xi})$ also has a limiting normal distribution with mean $g(\hat{\xi})$ and (asymptotic) variance:

$$Asy.var\left(g(\hat{\xi})\right) \equiv [g'(\xi)]^2 \bullet \theta(\xi)^2 . \qquad \text{(b6)}$$

The example of $g(\hat{\xi})$ we make use of here is $\widehat{PS}_M(\hat{\xi})$.

To obtain the gradients $g'(\xi)$, one makes use of Leibnitz's Rule (Bergin, 2015, p. 467; also available on the internet). In the case of the population share for the middle-income group,

$$g'(\xi) = f(1.5\xi)(1.5) - f(0.5\xi)(0.5)$$

so that

$$Asy.var(P\hat{S}_M) = [f(1.5\xi)(1.5) - f(0.5\xi)(0.5)]^2 \bullet [0.25 / f(\xi)^2] . \qquad \text{(b7)}$$

Hence the estimated asymptotic variance of $P\hat{S}_M$ is gotten by putting sample estimates into equation (b7).

B.1.2. Estimating with the Lognormal Distribution

Now to implement the formula in equation (b7) we need an expression for *f(•)*. A distribution-free approach for estimating this expression is provided by Davidson (2018) based on an integrated Epanechnikov kernel data smoothing technique. Kernel estimation for $f(\bullet)$ was also proposed by Lin, Wu and Ahmad (1980). While elegant, this approach is rather burdensome. It also requires a large sample size so that smoothing an empirical density function at a given point can be done reliably. We follow a simpler more direct approach of assuming that *f(•)* follows a specific functional form, namely a lognormal distribution. And again in relatively small data samples, imposing a function form restriction should reduce estimated standard errors on the distributional statistics. This seems a quite reasonable assumption when applied to income, earnings or wage distributions, but would not be advised for distributions with extremely long right-hand tails such as wealth distributions or distributions with a large number of negative observations. The estimates proposed here are thus not distribution-free, but for income and related distributions seem quite reasonable so as to allow a useful empirical tool box for distributional analysis.

So if *f(•)* follows a lognormal distribution of income *y*,

$$f(y) = \left[1 \,/\, \beta\sqrt{2\pi} \,\bullet y\right] \bullet exp[-(\ln y - \alpha)^2 \,/\, 2\beta^2] \quad for\ y \,\epsilon\, (0, +\infty) \qquad \text{(b8)}$$

where α is the mean of ln *y* and β^2 is the variance of ln *y*. The standard reference on the lognormal distribution is Aitchison and Brown (1957), but see also Kendall and Stuart (1969) and a lot of useful information on the distribution gleaned readily from the internet.

To implement the use of the lognormal distribution to calculate estimated standard errors, first compute the sample median income level $(\hat{\xi})$ and the sample mean and standard deviations of log incomes $(\hat{\alpha}, \hat{\beta})$ and the sample mean and standard deviation of incomes $(\hat{\mu}, \sigma)$.[2] Then compute the various cut-off bounds as functions of $\hat{\xi}$. If one plugs $\hat{\alpha}$ and $\hat{\beta}$ into equation (b8), these two parameter estimates fully specify *f(y)* as a function of the income variable *y*, referred to as $\hat{f}(y)$. Then evaluate $\hat{f}(y)$ at its sample median $\hat{\xi}$ and at the required cut-off values 0.5 $\hat{\xi}$ and 1.5 $\hat{\xi}$, referred to as $\hat{f}(\hat{\xi})$, $\hat{f}(0.5\hat{\xi})$, and $\hat{f}(1.5\hat{\xi})$, respectively. Hence,

$$Asy\hat{.}var(P\hat{S}_M) = \left[(1.5) \bullet \hat{f}(1.5\hat{\xi}) - (0.5) \bullet \hat{f}(0.5\hat{\xi})\right]^2 \bullet [0.25 \,/\, \hat{f}(\hat{\xi})^2]\,. \qquad \text{(b9)}$$

So the estimated standard error of $P\hat{S}_M$ is

$$S.E.(P\hat{S}_M) = \left[Asy\hat{.}var(P\hat{S}_M) \,/\, N\right]^{1/2}$$

$$= [(1.5) \bullet \hat{f}(1.5\hat{\xi}) - (0.5) \bullet \hat{f}(0.5\hat{\xi})] \bullet \left[(0.5) \,/\, \widehat{f}(\hat{\xi})\right] / \sqrt{N}\,. \qquad \text{(b10)}$$

[2] Given sample estimates of α and β^2 , one could generate implied estimates of μ and σ^2 (Kendall and Stuart, 1969, p. 68). But the linkages between them is highly nonlinear and we want to make use of the statistical properties of direct sample estimates of μ and σ^2 in the development of Appendices B and C. So we make use of direct sample estimates of both α, β^2 and μ, σ^2 .

In the case of weighted samples, estimates of ξ, α, β and μ and σ^2 should all be calculated in weighted fashion.

The standard error formulas for $\widehat{PS}$ over the lower and higher income ranges are obtained in similar fashion, and turn out to be simpler in form:

$$S.E.\left(P\hat{S}_L\right) = [(0.5) \bullet \hat{f}\left(0.5\hat{\xi}\right)] \bullet \left[(0.5) / \hat{f}\left(\hat{\xi}\right)\right] / \sqrt{N} \tag{b11}$$

$$S.E.\left(P\hat{S}_H\right) = [(1.5) \bullet \hat{f}\left(1.5\hat{\xi}\right)] \bullet \left[(0.5) / \hat{f}\left(\hat{\xi}\right)\right] / \sqrt{N}\,. \tag{b12}$$

The general approach of calculating standard errors of distributional tool box measures based on specifying an underyling income density function of the sample data – lognormal in the present case – and linking this density to sample quantile estimates is referred to as a *quantile function approach* since the integral expressions for the tool box measures are treated as functions of estimated quantiles. This is being forwarded as an alternative to the *cumulative moments approach* in Beach and Davidson (1983) which was motivated by methodological concerns. The new approach lends itself more readily to exploiting the relationships between the various tool box measures and to more accessible intuitive interpretation and easy programming/ calculation.

B.2 Standard Errors of Conditional Means and Population Shares

The quantile (or conditional) mean income levels can be written from first principles as

$$\mu_i = E(y \mid y \in i)$$

$$= \int_{R_i} y\, f(y)dy \;/\; \int_{R_i} f(y)dy \tag{b13}$$

where the integral takes over range R_i for income group *i* = *L, M, H*. So, for example, in the case of the middle-income group, *M*,

$$\mu_M = \int_{0.5\xi}^{1.5\xi} y\, f(y)dy \; / \; \int_{0.5\xi}^{1.5\xi} f(y)dy \, .$$

The numerator and denominator components of μ_i can be written separately as

$$\int_{R_i} y\, f(y)dy = N_i(\xi)$$

and $$\int_{R_i} f(y)dy = D_i(\xi)$$

where each is a function of the median ξ .

In order to facilitate looking at quantile mean income gaps later in this appendix, consider $\hat{\mu}_L$, $\hat{\mu}_M$, and $\hat{\mu}_H$ together. To do this, consider the six-element vector

$$g = (N_L,\ D_L,\ N_M,\ D_M,\ N_H,\ D_H)'$$

which is estimated by

$$\hat{g} = (\hat{N}_L,\ \hat{D}_L,\ \hat{N}_M,\ \hat{D}_M, \hat{N}_H,\ \hat{D}_H)'$$

where each element is calculated by summing up the income or observations (appropriately weighted if necessary) in each respective income group. To calculate appropriate standard errors, we proceed in several steps.

Step 1: The Asymptotic Joint Distribution of $\hat{N}_i$ and $\hat{D}_i$

We first need to establish the asymptotic distribution (and the asymptotic variances and covariances) of the separate components of $\hat{g}$. Note that each of the components of g is a continuous differentiable function of ξ, the median income level of the income distribution. Then by the multivariate Rao linkage theorem (Rao, 1965, p. 388) for continuous differentiable functions, the vector $\hat{g}$ is asymptotically joint normally distributed with (asymptotic) variance-covariance matrix

$$\underset{(6x6)}{\mathrm{W}} = \begin{bmatrix} Asy.var(\widehat{N}_L) & \cdots & Asy.cov(\widehat{N}_L, \widehat{D}_H) \\ \vdots & & \vdots \\ Asy.cov(\widehat{D}_H, \widehat{N}_L) & \cdots & Asy.var(\widehat{D}_H) \end{bmatrix} = [w_{ij}]$$

where $W = G\ \Theta\ G'$ (b14)

$$G = [g_1',\ g_2',\ g_3',\ g_4',\ g_5',\ g_6']'$$

with $g_1' = \frac{\partial N_L}{\partial \xi} = (0.5)(0.5\xi) \bullet f(0.5\xi)$

$$g_2' = \frac{\partial D_L}{\partial \xi} = (0.5)\ f(0.5\xi)$$

$$g_3' = \frac{\partial N_M}{\partial \xi} = (1.5)(1.5\xi) \bullet f(1.5\xi) - (0.5)(0.5\xi) \bullet f(0.5\xi)$$

$$g_4' = \frac{\partial D_M}{\partial \xi} = (1.5)\ f(1.5\xi) - (0.5)\ f(0.5\xi)$$

$$g_5' = \frac{\partial N_H}{\partial \xi} = -(1.5)(1.5\xi) \bullet f(1.5\xi)$$

$$g_6' = \frac{\partial D_H}{\partial \xi} = -(1.5)\ f(1.5\xi)\ .$$

Therefore,

$$\mathrm{W} = \Theta \bullet \begin{bmatrix} (g_1')^2 & \cdots & (g_1', g_6') \\ \vdots & & \vdots \\ (g_6', g_1') & \cdots & (g_6')^2 \end{bmatrix} = [w_{ij}] \qquad \text{(b15)}$$

where the scalar $\Theta = Asy.var(\hat{\xi}) = (0.5)(0.5)\ /\ [f(\xi)]^2$. So, for example,

$$Asy.var(\widehat{N}_L) = (g_1')^2 \bullet \Theta$$

$$= (0.5)^2\ (0.5\xi)^2 \bullet f(0.5\xi)^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right],$$

and $Asy.cov(\widehat{N}_L, \widehat{D}_L) = (g_1' g_2') \bullet \Theta$

$$= [(0.5)(0.5\xi) \bullet f(0.5\xi)] \bullet [(0.5) \bullet f(0.5\xi)] \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right],$$

which, as one would expect, is strictly positive.

Step 2: The Asymptotic Distribution of $\hat{\mu}_i$

A similar argument can be used to establish the (asymptotic) distributions of

$\hat{q}_i \equiv \hat{N}_i \,/\, \hat{D}_i \equiv \hat{\mu}_i$, $\qquad\qquad$ i = L, M, H

since simple division is a continuous differentiable function of the elements of $\hat{g}$. So again by Rao's linkage theorem, $\hat{q} = [\hat{q}_L,\ \hat{q}_M,\ \hat{q}_H]'$ is also asymptotically joint normally distributed with the (asymptotic) variance-covariance matrix

$$\underset{(3x3)}{Asy.var}(\hat{\mu}) \equiv Asy.var(\hat{q}) \equiv \underset{(3x3)}{V} = [v_{ij}]$$

$$= Q\,W\,Q' \qquad \text{(b16)}$$

where W is given in eq. (b15) and

$$Q = \begin{bmatrix} q_{11}, & q_{12}, & q_{13}, & q_{14}, & q_{15}, & q_{16} \\ q_{21}, & q_{22}, & q_{23}, & q_{24}, & q_{25}, & q_{26} \\ q_{31}, & q_{32}, & q_{33}, & q_{34}, & q_{35}, & q_{36} \end{bmatrix}$$

with

$$q_{11} = \frac{\partial q_L}{\partial N_L} = \frac{1}{D_L}$$

$$q_{12} = \frac{\partial q_L}{\partial D_L} = -\frac{N_L}{D_L^2}$$

$$q_{13} = \frac{\partial q_L}{\partial N_M} = 0$$

$$q_{14} = \frac{\partial q_L}{\partial D_M} = 0$$

$$q_{15} = \frac{\partial q_L}{\partial N_H} = 0$$

$$q_{16} = \frac{\partial q_L}{\partial D_H} = 0$$

$$q_{21} = \frac{\partial q_M}{\partial N_L} = 0$$

$$q_{22} = \frac{\partial q_M}{\partial D_L} = 0$$

$$q_{23} = \frac{\partial q_M}{\partial N_M} = \frac{1}{D_M}$$

$$q_{24} = \frac{\partial q_M}{\partial D_M} = -\frac{N_M}{D_M^2}$$

$$q_{25} = \frac{\partial q_M}{\partial N_H} = 0$$

$$q_{26} = \frac{\partial q_M}{\partial D_H} = 0$$

$$q_{31} = \frac{\partial q_H}{\partial N_L} = 0$$

$$q_{32} = \frac{\partial q_H}{\partial D_L} = 0$$

$$q_{33} = \frac{\partial q_H}{\partial N_M} = 0$$

$$q_{34} = \frac{\partial q_H}{\partial D_M} = 0$$

$$q_{35} = \frac{\partial q_H}{\partial N_H} = \frac{1}{D_H}$$

and $$q_{36} = \frac{\partial q_H}{\partial D_H} = -\frac{N_H}{D_H^2}.$$

Therefore,

$$Asy.var(\hat{\mu}_L) \equiv Asy.var(\hat{q}_L) = v_{11}$$

$$= (q_{11})^2\, w_{11} + (q_{12})^2\, w_{22} + 2(q_{11}q_{12})\, w_{12}$$

$$= \left(\frac{1}{D_L}\right)^2 \bullet Asy.var\left(\hat{N}_L\right) + \left(-\frac{N_L}{D_L^2}\right)^2 \bullet Asy.var\left(\hat{D}_L\right) \qquad \text{(b17)}$$

$$+ 2\left(\frac{1}{D_L}\right)\left(-\frac{N_L}{D_L^2}\right) \bullet Asy.cov\left(\hat{N}_L, \hat{D}_L\right).$$

The first term in (b17) captures the effect of variability in $\hat{N}_L$, the second term represents the variance effect of $\hat{D}_L$, and the third term picks us the covariance effect of $\hat{N}_L$ and $\hat{D}_L$ together.

Similarly,

$$Asy.var(\hat{\mu}_M) \equiv Asy.var(\hat{q}_M) = v_{22}$$

$$= (q_{23})^2\, w_{33} + (q_{24})^2\, w_{44} + 2(q_{23}q_{24})\, w_{34}$$

$$= \left(\frac{1}{D_M}\right)^2 \bullet Asy.var\left(\hat{N}_M\right) + \left(-\frac{N_M}{D_M^2}\right)^2 \bullet Asy.var\left(\hat{D}_M\right) \quad \text{(b18)}$$

$$+ 2\left(\frac{1}{D_M}\right)\left(-\frac{N_L}{D_M^2}\right) \bullet Asy.cov\left(\hat{N}_M, \hat{D}_M\right)$$

and

$$Asy.var(\hat{\mu}_H) \equiv Asy.var(\hat{q}_H) = v_{33}$$

$$= (q_{35})^2\, w_{55} + (q_{36})^2\, w_{66} + 2(q_{35}q_{36})\, w_{56}$$

$$= \left(\frac{1}{D_H}\right)^2 \bullet Asy.var\left(\hat{N}_H\right) + \left(-\frac{N_H}{D_H^2}\right)^2 \bullet Asy.var\left(\hat{D}_H\right) \quad \text{(b19)}$$

$$+ 2\left(\frac{1}{D_H}\right)\left(-\frac{N_H}{D_H^2}\right) \bullet Asy.cov\left(\hat{N}_H, \hat{D}_H\right).$$

But this derivation also allows one to calculate the full set of (asymptotic) covariances among the $\hat{\mu}_i$'s (whose results will be used below in section B.5).

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_M) = Asy.cov(\hat{q}_L, \hat{q}_M) = v_{12}$$

$$= (q_{11}\, q_{23})\, w_{13} + (q_{11}\, q_{24})\, w_{14} + (q_{12}q_{23})w_{23} + (q_{12}q_{24})w_{24}$$

$$= \left(\frac{1}{D_L}\right)\left(\frac{1}{D_M}\right) \bullet Asy.cov\left(\hat{N}_L, \hat{N}_M\right) + \left(\frac{1}{D_L}\right)\left(-\frac{N_M}{D_M^2}\right) \bullet Asy.cov\left(\hat{N}_L, \hat{D}_M\right) \quad \text{(b20)}$$

$$+ \left(-\frac{N_L}{D_L^2}\right)\left(\frac{1}{D_M}\right) \bullet Asy.cov\left(\hat{D}_L, \hat{N}_M\right) + \left(-\frac{N_L}{D_L^2}\right)\left(-\frac{N_M}{D_M^2}\right) \bullet Asy.cov\left(\hat{D}_L, \hat{D}_M\right).$$

Once again, the first term in (b20) captures the covariance effect between $\hat{N}_L$ and $\hat{N}_M$, the second term picks up that between $\hat{N}_L$ and $\hat{D}_M$, the third term represents the covariance effect between $\hat{D}_L$ and $\hat{N}_M$, and the last term reflects that between $\hat{D}_L$ and $\hat{D}_M$. Similarly,

$$Asy.cov(\hat{\mu}_M, \hat{\mu}_H) = Asy.cov(\hat{q}_H, \hat{q}_H) = v_{23}$$

$$= (q_{23}\, q_{35})\, w_{35} + (q_{23}\, q_{36})\, w_{36} + (q_{24}q_{35})w_{45} + (q_{24}q_{36})w_{46}$$

$$= \left(\frac{1}{D_M}\right)\left(\frac{1}{D_H}\right) \bullet Asy.cov\left(\hat{N}_M, \hat{N}_H\right) + \left(\frac{1}{D_M}\right)\left(-\frac{N_H}{D_H^2}\right) \bullet Asy.cov\left(\hat{N}_M, \hat{D}_H\right) \quad \text{(b21)}$$

$$+\left(-\frac{N_M}{D_M^2}\right)\left(\frac{1}{D_H}\right) \bullet Asy.cov\left(\widehat{D}_M, \widehat{N}_H\right) + \left(-\frac{N_M}{D_M^2}\right)\left(-\frac{N_H}{D_H^2}\right) \bullet Asy.cov\left(\widehat{D}_M, \widehat{D}_H\right)$$

and

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_H) = Asy.cov(\hat{q}_L, \hat{q}_H) = v_{13}$$

$$= (q_{11}\, q_{35})\, w_{15} + (q_{11}\, q_{36})\, w_{16} + (q_{12}q_{35})w_{25} + (q_{12}q_{36})w_{26}$$

$$= \left(\frac{1}{D_L}\right)\left(\frac{1}{D_H}\right) \bullet Asy.cov\left(\widehat{N}_L, \widehat{N}_H\right) + \left(\frac{1}{D_L}\right)\left(-\frac{N_H}{D_H^2}\right) \bullet Asy.cov\left(\widehat{N}_L, \widehat{D}_H\right) \qquad \text{(b22)}$$

$$+\left(-\frac{N_L}{D_L^2}\right)\left(\frac{1}{D_H}\right) \bullet Asy.cov\left(\widehat{D}_L, \widehat{N}_H\right) + \left(-\frac{N_L}{D_L^2}\right)\left(-\frac{N_H}{D_H^2}\right) \bullet Asy.cov\left(\widehat{D}_L, \widehat{D}_H\right) .$$

Consequently,

$$S.E.(\hat{\mu}_i) = \left[\frac{Asy\hat{.}var(\hat{\mu}_i)}{N}\right]^{1/2} \qquad \text{for } i = L, M, H. \qquad \text{(b23)}$$

One should note that the denominator terms D_i in these derivations are, in fact, the population shares or PS_i , one of our tool box measures. This general derivation has thus allowed one to establish that the $\widehat{PS}_i$'s are also asymptotically normally distributed with (asymptotic) variances given by w_{22} , w_{44} , and w_{66} , respectively, where

$$w_i = (g_i')^2 \bullet \theta \qquad \text{(b24)}$$

from eq. (b15), where here *i* = 2, 4, 6. Consequently,

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

$$= \left[\frac{(g_i')^2 \bullet \theta}{N}\right]^{1/2}$$

$$= |\, g_i'| \bullet (\theta \,/\, N)^{1/2} \qquad \text{for } i = L, M, H. \qquad \text{(b25)}$$

So, for example, in the case of the middle-income group, *M*,

$$Asy.var\left(\widehat{PS}_M\right) = (g_4')^2 \bullet \theta$$

$$= [(1.5)f(1.5\xi) - (0.5)f(0.5\xi)]^2 \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right]$$

which is what we earlier found in the first section of this appendix. Therefore,

$$S.E.(\widehat{PS}_M) = |(1.5)\hat{f}(1.5\hat{\xi}) - (0.5)\hat{f}(0.5\hat{\xi})| \left[\frac{0.5}{N^{1/2} \bullet \hat{f}(\hat{\xi})}\right]$$

where the hats indicate sample estimates of the unknowns.

One may note also that, since the D_i's are population shares, $\mu_i = N_i \, / \, PS_i$. Consequently, the (asymptotic) variance and covariance formulas in eq. (b17)-(b22) can be alternatively expressed in terms of the tool box measures μ_i and PS_i that the user would be calculating anyway. So, in the case of the (asymptotic) variances for the sample quantile or conditional means:

$$\frac{N_i}{D_i^2} = \frac{\mu_i}{PS_i} \qquad \text{and} \qquad \frac{1}{D_i} = \frac{1}{PS_i}$$

and

$$Asy.var(\hat{\mu}_L) = \left(\frac{1}{PS_L}\right)^2 \bullet Asy.var(\hat{N}_L) + \left(-\frac{\mu_L}{PS_L}\right)^2 \bullet Asy.var(\widehat{PS}_L) \qquad \text{(b26)}$$

$$+ 2\left(\frac{1}{PS_L}\right)\left(-\frac{\mu_L}{PS_L}\right) \bullet Asy.cov(\hat{N}_L, \hat{D}_L)$$

$$Asy.var(\hat{\mu}_M) = \left(\frac{1}{PS_M}\right)^2 \bullet Asy.var(\hat{N}_M) + \left(-\frac{\mu_M}{PS_M}\right)^2 \bullet Asy.var(\widehat{PS}_M) \qquad \text{(b27)}$$

$$+ 2\left(\frac{1}{PS_M}\right)\left(-\frac{\mu_M}{PS_M}\right) \bullet Asy.cov(\hat{N}_M, \hat{D}_M)$$

$$Asy.var(\hat{\mu}_H) = \left(\frac{1}{PS_H}\right)^2 \bullet Asy.var(\hat{N}_H) + \left(-\frac{\mu_H}{PS_H}\right)^2 \bullet Asy.var(\widehat{PS}_H) \qquad \text{(b28)}$$

$$+ 2\left(\frac{1}{PS_H}\right)\left(-\frac{\mu_H}{PS_H}\right) \bullet Asy.cov(\hat{N}_H, \hat{D}_H) \, .$$

B.3 Standard Errors of Income Shares

The income share of members of income group *i = L, M, H* in the distribution may be written from first principles as

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

In the case of the middle-income group *M*, then,

$$IS_M \ = \int_{0.5\xi}^{1.5\xi} \left(\frac{1}{\mu}\right) y\, f(y) dy \, .$$

The IS_i function can thus be seen to be a function of two parameters that are to be estimated from the sample data – ξ and μ , the median and the overall mean of the distribution of income. That is

$$IS_i = \ g_i(\xi, \mu)$$

which can be estimated by dividing total income of the members of group *i* by the total income receipts of all members of the data sample.

To address the jointedness of sample estimates of these two parameters, it is convenient to make use of results in Lin, Wu, and Ahmad (1980) – henceforth LWA – who establish that, under broad regularity conditions, $\hat{\xi}$ and $\hat{\mu}$ are asymptotically joint normally distributed with (asymptotic) variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

where $\sigma_{11} = Asy.var\left(\hat{\xi}\right) = \frac{(0.5)(0.5)}{[f(\xi)]^2}$

$$\sigma_{22} = Asy.var(\hat{\mu}) = \ \sigma^2$$

and $\sigma_{12} = Asy.cov\left(\hat{\xi},\ \hat{\mu}\right) = \frac{[\xi - \mu(1-0.5)]}{f(\xi)}$.

To apply these results, again recognize that $g_i(\xi, \mu)$ is a continuously differentiable function of its arguments and thence apply Rao’s multivariate linkage theorem to

$$g = \ [IS_L \, , IS_M \, , IS_H]' = \ [g_L \, , g_M \, , g_H]'$$

to establish that $\hat{g} = \left[\widehat{IS}_L, \widehat{IS}_M, \widehat{IS}_H\right]'$ are again asymptotically joint normally distributed with (asymptotic) variance-covariance matrix

$$\underset{(3x3)}{Asy.var}\left(\widehat{IS}\right) = Asy.var(\hat{g})$$

$$= G\,\Sigma\,G' \qquad \text{(b29)}$$

where Σ is given above and

$$G = \begin{bmatrix} g_{L1}, & g_{L2} \\ g_{M1}, & g_{M2} \\ g_{H1}, & g_{H2} \end{bmatrix}$$

where

$$g_{L1} = \frac{\partial g_L}{\partial \xi} = \left(\frac{1}{\mu}\right)(0.5)(0.5\xi) \bullet f(0.5\xi)$$

$$g_{L2} = \frac{\partial g_L}{\partial \mu} = -\left(\frac{1}{\mu}\right) IS_L$$

$$g_{M1} = \frac{\partial g_M}{\partial \xi} = \left(\frac{1}{\mu}\right)[(1.5)(1.5\xi) \bullet f(1.5\xi) - (0.5)(0.5\xi) \bullet f(0.5\xi)]$$

$$g_{M2} = \frac{\partial g_M}{\partial \xi} = -\left(\frac{1}{\mu}\right) IS_M$$

$$g_{H1} = \frac{\partial g_H}{\partial \xi} = -\left(\frac{1}{\mu}\right)(1.5)(1.5\xi) \bullet f(1.5\xi)$$

and $$g_{H2} = \frac{\partial g_H}{\partial \mu} = -\left(\frac{1}{\mu}\right) IS_H\,.$$

Therefore,

$$Asy.var\left(\widehat{IS}_L\right) = Asy.var(\hat{g}_L)$$

$$= (g_{L1})^2 \bullet \sigma_{11} + (g_{L2})^2 \bullet \sigma_{22} + 2(g_{L1}\,g_{L2}) \bullet \sigma_{12}$$

$$= \left(\frac{1}{\mu}\right)^2 [(0.5)(0.5\xi) \bullet f(0.5\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right] + \left(\frac{1}{\mu^2}\right) IS_L^2 \bullet \sigma^2 \qquad \text{(b30)}$$

$$+2\left[\left(\frac{1}{\mu}\right)(0.5)(0.5\xi) \bullet f(0.5\xi)\right]\left[\left(\frac{-1}{\mu}\right) IS_L\right]\left[\frac{\xi-(0.5)\mu}{f(\xi)}\right]$$

$$= \left(\frac{1}{\mu}\right)^2 [(0.5)(0.5\xi) \bullet f(0.5\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right] + \left(\frac{IS_L}{\mu}\right)^2 \bullet \sigma^2$$

$$-2\left(\frac{1}{\mu}\right)^2 [(0.5)(0.5\xi) \bullet f(0.5\xi)] \bullet IS_L \bullet \left[\frac{\xi-(0.5)\mu}{f(\xi)}\right],$$

$$Asy.var\left(\widehat{IS}_H\right) = Asy.var(\hat{g}_H)$$

$$= (g_{H1})^2 \bullet \sigma_{11} + (g_{H2})^2 \bullet \sigma_{22} + 2(g_{H1}\, g_{H2}) \bullet \sigma_{12}$$

$$= \left(\frac{1}{\mu}\right)^2 [(1.5)(1.5\xi) \bullet f(1.5\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right] + \left(\frac{1}{\mu}\right)^2 IS_H^2 \bullet \sigma^2 \qquad \text{(b31)}$$

$$+2\left(\frac{1}{\mu}\right)^2 [(1.5)(1.5\xi) \bullet f(1.5\xi)] \bullet IS_H \bullet \left[\frac{\xi-(0.5)\mu}{f(\xi)}\right],$$

and

$$Asy.var\left(\widehat{IS}_M\right) = Asy.var(\hat{g}_M)$$

$$= (g_{M1})^2 \bullet \sigma_{11} + (g_{M2})^2 \bullet \sigma_{22} + 2(g_{M1}\, g_{M2}) \bullet \sigma_{12}$$

$$= \left(\frac{1}{\mu}\right)^2 [(1.5)(1.5\xi) \bullet f(1.5\xi) - (0.5)(0.5\xi) \bullet f(0.5\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)}\right]$$

$$+ \left(\frac{1}{\mu}\right)^2 IS_M^2 \bullet \sigma^2 \qquad \text{(b32)}$$

$$-2\left(\frac{1}{\mu}\right)^2 |(1.5)(1.5\xi) \bullet f(1.5\xi) - (0.5)(0.5\xi) \bullet f(0.5\xi)|$$

$$\bullet IS_M \bullet \left[\frac{\xi-(0.5)\mu}{f(\xi)}\right].$$

In each case, the first term captures the effect of the variability in $\hat{\xi}$, the second term that of $\hat{\mu}$, and the third term picks up the joint covariance effect between $\hat{\xi}$ and $\hat{\mu}$.

Therefore,

$$S.E.\left(\widehat{IS}_i\right) = \left[\frac{Asy\hat{.}var(\widehat{IS}_i)}{N}\right]^{1/2} \qquad \text{(b33)}$$

where, again, unknowns are replaced by their sample estimates.

B.4 Standard Errors of Relative Mean Incomes

By relative mean income (RMI) is meant the ratio of conditional or quantile mean income to the overall mean income of an income distribution:

$$RMI_i = E(y \mid y \in i) \; / \, E(y) = \mu_i \, / \, \mu$$

for income group *i* = *L, M, H*. From first principles, then,

$$RMI_i = \left[\int_{R_i} y\, f(y)dy \; / \int_{R_i} f(y)dy\right] \; / \int_0^{\infty} y\, f(y)dy$$

$$= \frac{N_i(\xi) \,/\, D_i(\xi)}{\mu} \quad \text{in the notation of section B.2}$$

$$= \frac{N_i(\xi) \,/\, \mu}{D_i(\xi)}$$

$$= \frac{IS_i(\xi,\mu)}{PS_i(\xi)}. \qquad \text{(b34)}$$

So the relative mean income measure is identically equal to the ratio of the group's income share to its population share. So RMI_i can be estimated alternatively (and equivalently) as $\hat{\mu}_i \,/\, \hat{\mu}$ or as $\widehat{IS}_i \,/\, \widehat{PS}_i$. RMI_i can thus be viewed as the slope of the line segment of a Lorenz curve calculated over the range of the income group *i* of the distribution.

In order to calculate standard errors of $R\widehat{M}I_i$, one thus uses elements of the arguments already set out in the previous two sections. And again, we proceed in two steps.

Step 1: The Asymptotic Distribution of $\widehat{IS}_i$ and $\widehat{PS}_i$

Start by considering the six-element vector

$$g = (IS_L,\; PS_L,\; IS_M,\; PS_M,\; IS_H,\; PS_H)'$$

where recall that $PS_i \equiv D_i(\xi)$ from section B.2. The estimated vector then is

$$\hat{g} = (\widehat{IS}_L,\; \widehat{PS}_L,\; \widehat{IS}_M,\; \widehat{PS}_M, \widehat{IS}_H,\; \widehat{PS}_H)' .$$

Again, note that each of the components of g is a continuous differentiable function of ξ (and possibly μ). Rao's linkage theorem has already been used to establish the asymptotic distribution of estimates of the denominator components $PS_i = D_i(\xi)$ in section B.2, and Rao's linkage theorem was combined with the LWA results to establish the asymptotic distribution of estimates of the numerator components $IS_i = IS_i(\xi, \mu)$ in section B.3 above. We now combine these two sets of results by applying Rao's theorem to all the elements of vector *g* to establish that the components of $\hat{g}$ are again asymptotically joint normally distributed with (asymptotic) variance-covariance matrix

$$\underset{(6x6)}{\mathrm{W}} = \begin{bmatrix} Asy.var(\widehat{IS}_L) & \cdots & Asy.cov(\widehat{IS}_L, \widehat{PS}_H) \\ \vdots & & \vdots \\ Asy.cov(\widehat{PS}_H, \widehat{IS}_L) & \cdots & Asy.var(\widehat{PS}_H) \end{bmatrix}$$

where $W = G \, \Sigma \, G'$. (b35)

The (asymptotic) variance-covariance matrix for $\hat{\xi}$ and $\hat{\mu}$, Σ , is as defined in section B.3, and the matrix of partial derivatives, *G*, is given by

$$\underset{(6x2)}{G} = \begin{bmatrix} g_{L,11} & g_{L,12} \\ g_{L,21} & g_{L,22} \\ & \\ g_{M,11} & g_{M,12} \\ g_{M,21} & g_{M,22} \\ & \\ g_{H,11} & g_{H,12} \\ g_{H,21}, & g_{H,22} \end{bmatrix}$$

with $g_{i,11} = \frac{\partial IS_i}{\partial \xi}$ $\qquad g_{i,21} = \frac{\partial PS_i}{\partial \xi}$

$g_{i,12} = \frac{\partial IS_i}{\partial \mu}$ $\qquad g_{i,22} = \frac{\partial PS_i}{\partial \mu} = 0$

for *i* = *L, M, H*. Since this is a more general result than actually needed for our purposes, we focus just on the pairs $\widehat{IS}_L$ and $\widehat{PS}_L$, $\widehat{IS}_M$ and $\widehat{PS}_M$, and $\widehat{IS}_H$ and $\widehat{PS}_H$, each treated separately.

Note, incidentally, that this general result establishes the (asymptotic) joint normality of each of the pairs $\widehat{IS}_i$ and $\widehat{PS}_i$.

<u>Case of RMI_L :</u>

$$g_{L,11} = \frac{\partial IS_L}{\partial \xi} = \left(\frac{1}{\mu}\right)(0.5)(0.5\xi) \bullet f(0.5\xi)$$

$$g_{L,12} = \frac{\partial IS_L}{\partial \mu} = -\left(\frac{1}{\mu}\right) IS_L$$

$$g_{L,21} = \frac{\partial PS_L}{\partial \xi} = (0.5) \bullet f(0.5\xi)$$

$$g_{L,22} = \frac{\partial PS_L}{\partial \mu} = 0 \ .$$

Therefore, (asymptotic) variance-covariance matrix of $\widehat{IS}_L$ and $\widehat{PS}_L$ is given by

$$\begin{bmatrix} g_{L,11} & g_{L,12} \\ g_{L,21} & 0 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} g_{L,11} & g_{L,21} \\ g_{L,12} & 0 \end{bmatrix},$$

so that

$$Asy.var\left(\widehat{IS}_L\right) = \left(g_{L,11}\right)^2 \sigma_{11} + 2\left(g_{L,11}\, g_{L,12}\right) \sigma_{12} + \left(g_{L,12}\right)^2 \sigma^2$$

$$Asy.var\left(\widehat{PS}_L\right) = \left(g_{L,21}\right)^2 \sigma_{11}$$

and $$Asy.cov\left(\widehat{IS}_L\,, \widehat{PS}_L\right) = \left(g_{L,11}\, g_{L,21}\right) \sigma_{11} + \left(g_{L,12}\, g_{L,21}\right) \sigma_{12} \ .$$

The first two expressions are what we have found already; the covariance expression is what is new here.

<u>Case of RMI_H :</u>

$$g_{H,11} = \frac{\partial IS_H}{\partial \xi} = -\left(\frac{1}{\mu}\right)(1.5)(1.5\xi) \bullet f(1.5\xi)$$

$$g_{H,12} = \frac{\partial IS_H}{\partial \mu} = -\left(\frac{1}{\mu}\right) IS_H$$

$$g_{H,21} = \frac{\partial PS_H}{\partial \xi} = -(1.5) \bullet f(1.5\xi)$$

$$g_{H,22} = \frac{\partial PS_H}{\partial \mu} = 0 \ .$$

Therefore, (asymptotic) variance-covariance matrix of $\widehat{IS}_H$ and $\widehat{PS}_H$ is given by

$$\begin{bmatrix} g_{H,11} & g_{H,12} \\ g_{H,21} & 0 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} g_{H,11} & g_{H,21} \\ g_{H,12} & 0 \end{bmatrix},$$

so that

$$Asy.var(\widehat{IS}_H) = (g_{H,11})^2 \sigma_{11} + 2(g_{H,11}\, g_{H,12}) \sigma_{12} + (g_{H,12})^2 \sigma^2$$

$$Asy.var(\widehat{PS}_H) = (g_{H,21})^2 \sigma_{11}$$

and $\quad Asy.cov(\widehat{IS}_H , \widehat{PS}_H) = (g_{H,11}\, g_{H,21}) \sigma_{11} + (g_{H,12}\, g_{H,21}) \sigma_{12}$.

Case of RMI_M :

$$g_{M,11} = \frac{\partial IS_M}{\partial \xi} = \left(\frac{1}{\mu}\right) [(1.5)(1.5\xi) \bullet f(1.5\xi) - (0.5)(0.5\xi) \bullet f(0.5\xi)]$$

$$g_{M,12} = \frac{\partial IS_M}{\partial \mu} = -\left(\frac{1}{\mu}\right) IS_M$$

$$g_{M,21} = \frac{\partial PS_M}{\partial \xi} = (1.5) \bullet f(1.5\xi) - (0.5) \bullet f(0.5\xi)$$

$$g_{M,22} = \frac{\partial PS_M}{\partial \mu} = 0 \ .$$

Therefore, the (asymptotic) variance-covariance matrix of $\widehat{IS}_M$ and $\widehat{PS}_M$ is again given by

$$\begin{bmatrix} g_{M,11} & g_{M,12} \\ g_{M,21} & 0 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} g_{M,11} & g_{M,21} \\ g_{M,12} & 0 \end{bmatrix},$$

so that

$$Asy.var(\widehat{IS}_M) = (g_{M,11})^2 \sigma_{11} + 2(g_{M,11}\, g_{M,12}) \sigma_{12} + (g_{M,12})^2 \sigma^2$$

$$Asy.var(\widehat{PS}_M) = (g_{M,21})^2 \sigma_{11}$$

and $\quad Asy.cov(\widehat{IS}_M, \widehat{PS}_M) = (g_{M,11}\, g_{M,21})\, \sigma_{11} + (g_{M,12}\, g_{M,21})\, \sigma_{12}$.

Step 2: The Asymptotic Distribution of $R\widehat{M}I_i$

The three cases can now all be considered together. What we are working towards is the (asymptotic) distribution of $R\widehat{M}I_i$. So consider the ratio

$$RMI_i \equiv q_i = \frac{IS_i}{PS_i} \qquad \text{and its estimate } \hat{q}_i \equiv R\widehat{M}I_i = \frac{I\hat{S}_i}{P\hat{S}_i},$$

and the corresponding (asymptotic) variance-covariance matrix

$$\underset{(2x2)}{\Lambda_i} = \begin{bmatrix} Asy.var(\widehat{IS}_i) & Asy.cov(\widehat{IS}_i, \widehat{PS}_i) \\ Asy.cov(\widehat{PS}_i, \widehat{IS}_i) & Asy.var(\widehat{PS}_i) \end{bmatrix} = [\lambda_{ij}].$$

Again, by using the Rao linkage theorem for continuous differentiable functions, one can establish that $R\widehat{M}I_i = \hat{q}_i$ is asymptotically normally distributed with (asymptotic) variance

$$Asy.var(R\widehat{M}I_i) = Q_i' \Lambda_i Q_i \tag{b36}$$

with $\quad Q_i = [q_1, q_2]'$ where

$$q_1 = \frac{\partial q}{\partial IS_i} = \frac{1}{PS_i}$$

and $\quad q_2 = \frac{\partial q}{\partial PS_i} = -\frac{IS_i}{PS_i^2}$.

Therefore,

$$Asy.var(R\widehat{M}I_i) = \left(\frac{1}{PS_i}\right)^2 \bullet Asy.var(\widehat{IS}_i) + \left(-\frac{IS_i}{PS_i^2}\right)^2 \bullet Asy.var(\widehat{PS}_i)$$

$$+ 2\left[\left(\frac{1}{PS_i}\right)\left(-\frac{IS_i}{PS_i^2}\right)\right] \bullet Asy.cov(\widehat{IS}_i, \widehat{PS}_i)$$

$$= \left(\frac{1}{PS_i}\right)^2 \bullet Asy.var(\widehat{IS}_i) + \left(\frac{RMI_i}{PS_i}\right)^2 \bullet Asy.var(\widehat{PS}_i)$$

$$- 2\left(\frac{RMI_i}{PS_i^2}\right) \bullet Asy.cov\left(\widehat{IS}_i, \widehat{PS}_i\right) \qquad \text{(b37)}$$

for *i* = *L, M, H.*

Consequently,

$$S.E.\left(R\widehat{M}I_i\right) = S.E.(\hat{\mu}_i / \hat{\mu}) = \left[\frac{Asy.var(R\widehat{M}I_i)}{N}\right]^{1/2} . \qquad \text{(b38)}$$

B.5 Standard Errors of Quantile Mean Income Gaps

By conditional or quantile mean income gaps is meant the difference in mean income levels of two different income groups within the income distribution:

$$\mu_i - \mu_j \equiv E(y \mid y \epsilon i) - E(y \mid y \epsilon j) \quad \text{where} \quad i \neq j$$

and *i, j* = *L, M, or H.*

It turns out that the lengthy development of section B.2 now provides simple answers for the standard errors of the estimated quantile mean income gaps $\hat{\mu}_i - \hat{\mu}_j$ for $i \neq j$. Since this gap is the difference between quantile mean statistics, it follows that

$$Asy.var\left(\hat{\mu}_i - \hat{\mu}_j\right) = Asy.var(\hat{\mu}_i) - 2Asy.cov\left(\hat{\mu}_i, \hat{\mu}_j\right) + Asy.var\left(\hat{\mu}_j\right) \qquad \text{(b39)}$$

where the (asymptotic) variances are given by eq. (b26)-(b28) and the (asymptotic) covariances by eq. (b20)-(b22). The covariances may also be written in more convenient form as:

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_M) = \left(\frac{1}{PS_L}\right)\left(\frac{1}{PS_M}\right) \bullet Asy.cov\left(\widehat{N}_L, \widehat{N}_M\right)$$

$$- \left(\frac{1}{PS_L}\right)\left(\frac{\mu_M}{PS_M}\right) \bullet Asy.cov\left(\widehat{N}_L, \widehat{D}_M\right) \qquad \text{(b40)}$$

$$- \left(\frac{1}{PS_M}\right)\left(\frac{\mu_L}{PS_L}\right) \bullet Asy.cov\left(\widehat{D}_L, \widehat{N}_M\right)$$

$$+ \left(\frac{\mu_L}{PS_L}\right)\left(\frac{\mu_M}{PS_M}\right) \bullet Asy.cov\left(\widehat{D}_L, \widehat{D}_M\right) ,$$

$$Asy.cov(\hat{\mu}_M, \hat{\mu}_H) = \left(\frac{1}{PS_M}\right)\left(\frac{1}{PS_H}\right) \bullet Asy.cov(\hat{N}_M, \hat{N}_H)$$

$$- \left(\frac{1}{PS_M}\right)\left(\frac{\mu_H}{PS_H}\right) \bullet Asy.cov(\hat{N}_M, \hat{D}_H) \tag{b41}$$

$$- \left(\frac{1}{PS_H}\right)\left(\frac{\mu_M}{PS_M}\right) \bullet Asy.cov(\hat{D}_M, \hat{N}_H)$$

$$+ \left(\frac{\mu_M}{PS_M}\right)\left(\frac{\mu_H}{PS_H}\right) \bullet Asy.cov(\hat{D}_M, \hat{D}_H) \ ,$$

and

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_H) = \left(\frac{1}{PS_L}\right)\left(\frac{1}{PS_H}\right) \bullet Asy.cov(\hat{N}_L, \hat{N}_H)$$

$$- \left(\frac{1}{PS_L}\right)\left(\frac{\mu_H}{PS_H}\right) \bullet Asy.cov(\hat{N}_L, \hat{D}_H) \tag{b42}$$

$$- \left(\frac{1}{PS_H}\right)\left(\frac{\mu_L}{PS_L}\right) \bullet Asy.cov(\hat{D}_L, \hat{N}_H)$$

$$+ \left(\frac{\mu_L}{PS_L}\right)\left(\frac{\mu_H}{PS_H}\right) \bullet Asy.cov(\hat{D}_L, \hat{D}_H) \ .$$

Consequently, it follows that

$$S.E.(\hat{\mu}_i - \hat{\mu}_j) = \left[\frac{Asy.\hat{v}ar(\hat{\mu}_i - \hat{\mu}_j)}{N}\right]^{1/2} \tag{b43}$$

where all unknowns are replaced by their sample estimates.

And, while one is at it, one could also calculate the standard error for the relative or proportional income gap or differential

$$\hat{q} = \frac{(\hat{\mu}_i - \hat{\mu}_j)}{\hat{\mu}_j} = \frac{\hat{\mu}_i}{\hat{\mu}_j} - 1 \qquad \text{for } i > j \ .$$

Again by Rao's linkage theorem, $\hat{q}$ will also be asymptotically normally distributed with (asymptotic) variance

$$Asy.var(\hat{q}) = \left(\frac{1}{\mu_j}\right)^2 \bullet Asy.var(\hat{\mu}_i) + \left(\frac{\mu_i}{\mu_j^2}\right)^2 \bullet Asy.var(\hat{\mu}_j)$$

$$- 2\left(\frac{1}{\mu_j}\right)\left(\frac{\mu_i}{\mu_j^2}\right) \bullet Asy.cov(\hat{\mu}_i, \hat{\mu}_j) \tag{b44}$$

for $i > j$. Thus the standard error of the relative income gap is

$$S.E.(\hat{q}) = \left[\frac{Asy\hat{.}var(\hat{q})}{N}\right]^{1/2} \quad \text{(b45)}$$

as well.

B.6 Standard Error of an Income Polarization Measure and a Measure of Skewness

One could also consider a measure of *polarization* in the distribution of income as simply the sum of the lower and higher population shares:

$$P\hat{O}L = P\hat{S}_L + P\hat{S}_H .$$

Consequently, it follows that

$$Asy.var(P\hat{O}L) = Asy.var(P\hat{S}_L) + Asy.var(P\hat{S}_H) + 2\,Asy.cov(P\hat{S}_L, P\hat{S}_H). \quad \text{(b46)}$$

The two (asymptotic) variances have already been derived in section B.2, eq. (b15), as

$$Asy.var(\widehat{PS}_L) = [(0.5) \bullet f(0.5\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right]$$

and $$Asy.var(\widehat{PS}_H) = [-(1.5) \bullet f(1.5\xi)]^2 \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right].$$

It also can be seen that the (asymptotic) covariance is also available from eq. (b21) as the term w_{26} :

$$Asy.var(\widehat{PS}_L, \widehat{PS}_H) = [(0.5) \bullet f(0.5\xi)][-(1.5) \bullet f(1.5\xi)] \bullet \left[\frac{(0.5)(0.5)}{f(\xi)^2}\right] \quad \text{(b47)}$$

which, not surprisingly, turns out to be negative. Also, as a linear function of two (asymptotically) normal random variables, $P\hat{O}L$ is (asymptotically) normally distributed as well. Consequently, from (b46) above, it follows that

$$S.E.(P\hat{O}L) = \left[\frac{Asy\hat{.}var(P\hat{O}L)}{N}\right]^{1/2} . \qquad \text{(b48)}$$

Finally, a measure of skewness may be a useful component of the tool box of distributional measures. Skewness of a distribution is the degree to which a distribution is asymmetric around its midpoint. A normal distribution is symmetric about its mean. In the case of a distribution of incomes, however, such distributions are distinctly non-symmetric because of a very long right-hand tail representing a very small proportion of recipients with very high incomes while the majority of recipients have incomes below the mean. Consequently, one measure of skewness of a distribution is the ratio of the median or "typical" or "middle-most" income level to the overall mean of the distribution. For a symmetric distribution, the ratio is one. The more extreme the right-hand tail, the larger the gap between the mean and median and the lower the value of the median-to-mean ratio, $\hat{\xi} / \hat{\mu}$, below the value of one.

It would thus be useful to be able to estimate and do standard statistical inference on the median-to-mean ratio from sample data. It turns out that the arguments above (in section B.3) allow one to do so. Specifically, the work of Lin, Wu and Ahmad (1980) establishes that the skewness ratio $\hat{\xi} / \hat{\mu}$ is indeed asymptotically normally distributed with (asymptotic) variance

$$Asy.var(\hat{\xi} / \hat{\mu}) = \left[\left(\frac{\xi}{\mu}\right)^2 \sigma^2 - 2\left(\frac{\xi}{\mu}\right) \sigma_{12} + \sigma_{11}\right] / \ \mu^2 \qquad \text{(b49)}$$

where

$$\sigma_{11} = Asy.var(\hat{\xi}) = \frac{(0.5)(0.5)}{[f(\xi)^2]}$$

$$\sigma_{12} = Asy.cov(\hat{\xi},\ \hat{\mu}) = \frac{[\xi - \mu(1-0.5)}{f(\xi)}$$

and $\sigma^2 = Asy.cov(\hat{\mu})$.

Consequently,

$$S.E.\left(\hat{\xi} \,/\, \hat{\mu}\right) = \left[\frac{Asy\hat{.}var\left(\hat{\xi} \,/\, \hat{\mu}\right)}{N}\right]^{1/2} \tag{b50}$$

with sample estimates replacing all unknowns.

Technical Appendix C

A Blended Approach to Standard Errors for Percentile Cut-Offs and Percentile-Based Statistics

The idea of using a quantile function approach to calculate standard errors of disaggregative distributional statistics can readily apply to measures other than median-based statistics. Appendix A provided one approach to calculating standard errors of various percentile-based measures. But the quantile function approach of Appendix B can be applied here as well. The result is a blended approach to calculating standard errors for percentile-based statistics. As it turns out, in the case of some tool box measures, the resulting standard errors formulas are distribution-free.

C.1 Standard Errors of Percentile Cut-Offs and Percentile Income Gaps

To illustrate, assume, as in Appendix A, that there are K (ordered) percentile income groups indexed by $i = 1, \ldots, K$. For deciles, $K = 10$. Again, let p_i represent the (cumulative) proportion of the ordered income groups. For deciles, $p_1 = 0.1, p_2 = 0.2, \ldots, p_{10} = 1.0$. And let income cut-offs or quantiles that divide the different income groups be $\xi_1, \xi_2, \ldots, \xi_{K-1}$. Also let $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_{K-1})'$ be a vector of K-1 sample quantile estimates. Then it is well established in the statistics literature (for example, Wilks, 1962, p. 274) that, if $\hat{\xi}$ is a vector of K-1 sample quantiles from a random sample of size N drawn from a continuous population density $f(\bullet)$ such that the ξ_i's are uniquely defined and if $f_i \equiv f(\xi_i) > 0$ for all $i = 1, \ldots, K$-1, then the vector

$\sqrt{N}\ (\hat{\xi} - \xi)$ converges in distribution to a *K-1* variate normal distribution with mean zero and covariance matrix

$$\Lambda = \begin{bmatrix} \frac{p_1(1-p_1)}{f_1^2} & \cdots & \frac{p_1(1-p_{K-1})}{f_1 \bullet f_{K-1}} \\ \vdots & & \vdots \\ \frac{p_1(1-p_{K-1})}{f_1 \bullet f_{K-1}} & \cdots & \frac{p_{K-1}(1-p_{K-1})}{f_{K-1}^2} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1,K-1} \\ \vdots & \ddots & \vdots \\ \lambda_{1,K-1} & \cdots & \lambda_{K-1,K-1} \end{bmatrix} \qquad \text{(c1)}$$

The median, ξ_5 , is only a special case of this more general result.

Now each of these (asymptotic) variances and covariances is clearly not distribution-free as the denominator in each such term involves $f(\xi_i)$. But the quantile function approach developed in Appendix B is forwarded to deal with just this problem. As set out in section B.1.2, one can calculate $\hat{f}(\hat{\xi}_i)$ for any *i* based on, say, the estimated lognormal distribution. Hence

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

$$= \left[p_i(1 - p_i) \;/\; N \bullet \hat{f}(\hat{\xi}_i)^2\right]^{1/2}. \qquad \text{(c2)}$$

One can similarly use this approach to calculate estimated standard errors for the percentile income gaps

$\xi_j - \xi_i$ for $j \neq i$:

$$Asy.var(\hat{\xi}_j - \hat{\xi}_i) = Asy.var(\hat{\xi}_j) - 2Asy.cov(\hat{\xi}_i, \hat{\xi}_j) + Asy.var(\hat{\xi}_i) \qquad \text{(c3)}$$

where each of these terms can be taken from the variance-covariance matrix Λ in eq. (c1). As a result,

$$S.E.(\hat{\xi}_j - \hat{\xi}_i) = \left[\frac{Asy\hat{.}var(\hat{\xi}_j - \hat{\xi}_i)}{N}\right]^{1/2} \qquad \text{(c4)}$$

where again the denominator in each of the terms in (c3) is estimated by the lognormal estimated density.

Similarly, in the case of relative or proportional income gaps or differentials

$$q = (\xi_j - \xi_i) / \xi_i = \left(\frac{\xi_j}{\xi_i}\right) - 1 \qquad \text{for } j > i ,$$

$$Asy.var(\hat{q}) = \left(\frac{\xi_j}{\xi_i^2}\right)^2 \bullet Asy.var(\hat{\xi}_i) - 2\left(\frac{1}{\xi_i}\right)\left(\frac{\xi_j}{\xi_i^2}\right) \bullet Asy.cov(\hat{\xi}_i, \hat{\xi}_j)$$

$$+ \left(\frac{1}{\xi_i}\right)^2 \bullet Asy.var(\hat{\xi}_j) . \qquad \text{(c5)}$$

Consequently,

$$S.E.(\hat{q}) = \left[\frac{Asy.\hat{v}ar(\hat{q})}{N}\right]^{1/2} . \qquad \text{(c6)}$$

And all unknowns in (c5) are estimated by the estimated lognormal density.

C.2 Standard Errors of Quantile Means

One can make use of some of the key relationships from Appendix B here as well. Again, divide the income distribution in three income groups: the lower 20 percent (or L group), the middle 60 percent (or M group), and top 20 percent (or H group). The cut-off are then the two percentiles $p_2 = 0.2$ and $p_8 = 0.8$. More generally, it is notationally convenient to refer to the two percentiles as p_a and p_b , where $0 < p_a < p_b < 1$ and the corresponding quantile income cut-off values as $\xi_a < \xi_b$.

The quantile mean can be written

$$\mu_i = \int_{R_i} y\, f(y) dy \;/\; \int_{R_i} f(y) dy \qquad \text{for } i = L, M, H . \qquad \text{(c7)}$$

For the lower income group, R_L runs from 0 to ξ_a , R_M for the middle group runs from ξ_a to ξ_b , and for the higher income group R_H runs from ξ_b up. Since the percentile shares are given by p_a and p_b , the denominator in (c7) can be written as

$$D_L = p_a \qquad \text{for } i = L$$

$$D_M = p_b - p_a \qquad \text{for } i = M$$

and $D_H = 1 - p_b$ for $i = H$.

The numerator in (c7) can be expressed as

$$N_L(\xi_a, \xi_b) = \int_0^{\xi_a} y\, f(y) dy \qquad \text{for } i = L$$

$$N_M(\xi_a, \xi_b) = \int_{\xi_a}^{\xi_b} y\, f(y) dy \qquad \text{for } i = M$$

and $$N_H(\xi_a, \xi_b) = \int_{\xi_a}^{\infty} y\, f(y) dy \qquad \text{for } i = H .$$

Thus $$\mu_i = N_i(\xi_a, \xi_b) \,/\, D_i \qquad \text{for } i = L, M, H . \qquad \text{(c8)}$$

In order to facilitate looking at quantile mean income gaps later in this appendix, consider the sample mean estimates $\hat{\mu}_L$, $\hat{\mu}_M$, and $\hat{\mu}_H$ together. So consider the three-element vector

$$m = (\mu_L\, , \mu_M\, , \mu_H)'$$

$$= \left[\frac{N_L(\xi_a,\xi_b)}{D_L}, \frac{N_M(\xi_a,\xi_b)}{D_M}, \frac{N_H(\xi_a,\xi_b)}{D_H}\right]'$$

which is estimated directly as $\hat{m} = (\hat{\mu}_L\, , \hat{\mu}_M\, , \hat{\mu}_H)'$. So the terms in m are continuous differentiable functions of ξ_a and ξ_b whose estimates have the asymptotic normal distribution given in the previous section C.1. For a well behaved underlying income density function $f(\bullet)$, a multivariate version of the Rao linkage theorem (Rao, 1965, p. 388) shows that $\hat{m}$ is also asymptotically joint normally distributed with the (asymptotic) 3x3 variance-covariance matrix

$$Asy.var(\hat{m}) \equiv V = G' \Lambda G \qquad \text{(c9)}$$

where $\Lambda = [\lambda_{ij}]$ is the 2x2 (asymptotic) variance-covariance matrix of $(\hat{\xi}_a\, , \hat{\xi}_b)'$ given above in (c1), and where

$$G = \begin{bmatrix} g_{aL} & g_{aM} & g_{aH} \\ g_{bL} & g_{bM} & g_{bH} \end{bmatrix}$$

with $$g_{aL} = \frac{\partial \mu_L}{\partial \xi_a} = \xi_a \bullet f(\xi_a) \,/\, D_L$$

$$g_{bL} = \frac{\partial \mu_L}{\partial \xi_b} = 0$$

$$g_{aM} = \frac{\partial \mu_M}{\partial \xi_a} = -\xi_a \bullet f(\xi_a) \; / \; D_M \qquad \text{(c10)}$$

$$g_{bM} = \frac{\partial \mu_M}{\partial \xi_b} = \xi_b \bullet f(\xi_b) \; / \; D_M$$

$$g_{aH} = \frac{\partial \mu_H}{\partial \xi_a} = 0$$

$$g_{bH} = \frac{\partial \mu_H}{\partial \xi_b} = -\xi_b \bullet f(\xi_b) \; / \; D_H \; .$$

Working out the calculations, then,

$$Asy.var(\hat{\mu}_L) = \frac{\xi_a^2 \bullet f(\xi_a)^2}{D_L^2} \bullet \lambda_{aa}$$

$$= \frac{\xi_a^2 \bullet f(\xi_a)^2}{p_a^2} \left[\frac{p_a(1-p_a)}{f(\xi_a)^2}\right]$$

$$= \left(\frac{1-p_a}{p_a}\right) \bullet \xi_a^2 \, , \qquad \text{(c11)}$$

$$Asy.var(\hat{\mu}_H) = \frac{\xi_b^2 \bullet f(\xi_b)^2}{D_H^2} \bullet \lambda_{bb}$$

$$= \frac{\xi_b^2 \bullet f(\xi_b)^2}{(1-p_b)^2} \left[\frac{p_b(1-p_b)}{f(\xi_b)^2}\right]$$

$$= \left(\frac{p_b}{1-p_b}\right) \bullet \xi_b^2 \, , \qquad \text{(c12)}$$

$$Asy.var(\hat{\mu}_M) = \left(\frac{1}{D_M^2}\right) [\xi_a^2 \bullet f(\xi_a)^2 \bullet \lambda_{aa} + \xi_b^2 \bullet f(\xi_b)^2 \bullet \lambda_{bb}$$

$$-2\, \xi_a\, \xi_b \bullet f(\xi_a) f(\xi_b) \bullet \lambda_{ab}]$$

$$= \left(\frac{1}{p_b - p_a}\right)^2 \bullet [p_a(1-p_a) \bullet \xi_a^2 + p_b(1-p_b) \bullet \xi_b^2 \qquad \text{(c13)}$$

$$-\, 2\, p_a(1-p_b) \bullet \xi_a\, \xi_b] \, .$$

Covariance results can also be obtained to be used below in section C.5.

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_M) = [g_{aL}, g_{bL}] \begin{bmatrix} \lambda_{aa} & \lambda_{ab} \\ \lambda_{ba} & \lambda_{bb} \end{bmatrix} \begin{bmatrix} g_{aM} \\ g_{bM} \end{bmatrix}$$

$$= -\frac{\xi_a^2 \bullet f(\xi_a)^2}{D_L \bullet D_M} \bullet \lambda_{aa} + \frac{\xi_a\, \xi_b \bullet f(\xi_a) f(\xi_b)}{D_L \bullet D_M} \bullet \lambda_{ab}$$

$$= \frac{\xi_a}{(p_b - p_a)} \left[(1 - p_b)\xi_b - (1 - p_a)\xi_a\right], \qquad \text{(c14)}$$

$$Asy.cov(\hat{\mu}_M, \hat{\mu}_H) = [g_{aM}, g_{bM}] \begin{bmatrix} \lambda_{aa} & \lambda_{ab} \\ \lambda_{ba} & \lambda_{bb} \end{bmatrix} \begin{bmatrix} g_{aH} \\ g_{bH} \end{bmatrix}$$

$$= -\frac{\xi_b^2 \bullet f(\xi_b)^2}{D_M \bullet D_H} \bullet \lambda_{bb} + \frac{\xi_a \, \xi_b \bullet f(\xi_a) f(\xi_b)}{D_M \bullet D_H} \bullet \lambda_{ab}$$

$$= \frac{\xi_b}{(p_b - p_a)} \left[p_a \, \xi_a - p_b \, \xi_b\right], \qquad \text{(c15)}$$

$$Asy.cov(\hat{\mu}_L, \hat{\mu}_H) = [g_{aL}, g_{bL}] \begin{bmatrix} \lambda_{aa} & \lambda_{ab} \\ \lambda_{ba} & \lambda_{bb} \end{bmatrix} \begin{bmatrix} g_{aH} \\ g_{bH} \end{bmatrix}$$

$$= -\frac{\xi_a \, \xi_b \bullet f(\xi_a) f(\xi_b)}{D_L \bullet D_H} \bullet \lambda_{ab}$$

$$= -\xi_a \bullet \xi_b \, . \qquad \text{(c16)}$$

So, following from the above asymptotic variance results, it can be seen that

$$S.E.(\hat{\mu}_i) = \left[\frac{Asy\hat{.}var(\hat{\mu}_i)}{N}\right]^{1/2} \qquad \text{for } i = L, M, H\,. \qquad \text{(c17)}$$

C.3 Standard Errors of Relative-Mean Incomes

The relative-mean income is defined as

$$RMI_i = E(y \mid y \in i) \; / \; E(y) = \mu_i \, / \, \mu \qquad \text{for } i = L, M, H\,.$$

This then involves linking the (asymptotic) distribution of the $\hat{\mu}_i$'s with that of $\hat{\mu}$, the overall sample mean. One can go about this in either of two alternative ways.

C.3.1 The Adding-Up Approach

In the first or adding-up approach, one recognizes that

$$\mu = D_L \mu_L + D_M \mu_M + D_H \mu_H \qquad \text{(c18)}$$

and similarly,

$$\hat{\mu} = D_L \hat{\mu}_L + D_M \hat{\mu}_M + D_H \hat{\mu}_H \,. \qquad \text{(c19)}$$

Since $\hat{\mu}$ is a linear function of (asymptotically) normal random variables, so also $\hat{\mu}$ is (asymptotically) jointly normally distributed (along with the $\hat{\mu}_i$'s) as well.

Now $\quad Cov(\hat{\mu}_i, \hat{\mu}) \equiv E[(\hat{\mu}_i - \mu_i)(\hat{\mu} - \mu)]$.

Substituting in the expressions for μ and $\hat{\mu}$ from (c18) and (c19) leads to

$$Cov(\hat{\mu}_i, \hat{\mu}) = D_i \bullet Var(\hat{\mu}_i) + \sum_{j \neq i} D_j \bullet Cov(\hat{\mu}_i, \hat{\mu}_j) \,.$$

For example, when $i = M$,

$$Cov(\hat{\mu}_M, \hat{\mu}) = D_M \bullet Var(\hat{\mu}_M) + D_L \bullet Cov(\hat{\mu}_M, \hat{\mu}_L) + D_H \bullet Cov(\hat{\mu}_M, \hat{\mu}_H) \,.$$

Since this is exact for all N, it also holds asymptotically as

$$Asy.cov(\hat{\mu}_i, \hat{\mu}) = D_i \bullet Asy.var(\hat{\mu}_i) + \sum_{j \neq i} D_j \bullet Asy.cov(\hat{\mu}_i, \hat{\mu}_j) \,. \qquad \text{(c20)}$$

Then, since $\hat{\mu}_i$ and $\hat{\mu}$ are asymptotically joint normally distributed with known variance-covariance structure established in the previous section along with (c20), one can again make use of the Rao linkage theorem to further establish that, if

$$RMI_i = \mu_i \,/\, \mu \quad \text{and} \quad R\widehat{M}I_i = \hat{\mu}_i \,/\, \hat{\mu} \qquad \text{for } i = L, M, H \,,$$

then $\quad Asy.var\big(R\widehat{M}I_i\big) = Q' \, W \, Q$

where $W = \begin{bmatrix} Asy.var(\hat{\mu}_i) & Asy.cov(\hat{\mu}_i, \hat{\mu}) \\ Asy.cov(\hat{\mu}, \hat{\mu}_i) & Asy.var(\hat{\mu}_i) \end{bmatrix}$

and $\quad Q = [q_1, q_2]'$ with

$$q_1 = \frac{\partial RMI_i}{\partial \mu_i} = \frac{1}{\mu} \quad \text{and} \quad q_2 = \frac{\partial RMI_i}{\partial \mu} = -\frac{\mu_i}{\mu^2} \,.$$

Consequently, for $i = L, M, H,$

$$Asy.var\big(R\widehat{M}I_i\big) = \left(\frac{1}{\mu^2}\right) \bullet Asy.var(\hat{\mu}_i) + \left(\frac{\mu_i}{\mu^2}\right)^2 \bullet Asy.var(\hat{\mu})$$
$$- 2\left(\frac{1}{\mu}\right)\left(\frac{\mu_i}{\mu^2}\right) \bullet Asy.cov(\hat{\mu}_i, \hat{\mu}) \qquad \text{(c21)}$$

where $Asy.var(\hat{\mu}) = \sigma^2$, and correspondingly

$$S.E.(R\widehat{M}I_i) = S.E.(\hat{\mu}_i / \hat{\mu}) = \left[\frac{Asy\hat{.}var(R\widehat{M}I_i)}{N}\right]^{1/2} \quad \text{(c22)}$$

for all *i* = *L, M, H*, where all unknowns are replaced by their sample estimates. (This adding-up approach was not available in Appendix B because the D_i's were not given and had to be estimated, so $\hat{\mu}$ could not be expressed as a simple linear function of random variables.)

C.3.2 The Joint Distribution Approach

An alternative and perhaps more direct approach is to explicitly incorporate the joint randomness of the $\hat{\mu}_i$'s and the sample mean $\hat{\mu}$. To do this, one can again make use of the results of Lin, Wu and Ahmad (1980) – henceforth LWA.

The relative-mean income ratio can be written as

$$RMI_i \equiv \frac{\mu_i}{\mu} = \left[\int_{R_i} \frac{1}{\mu} y\, f(y)dy \;/\; \int_{R_i} f(y)dy\right]$$

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

$$= N_i(\xi_a, \xi_b, \mu) / D_i \qquad \text{for } i = L, M, H \quad \text{(c23)}$$

where D_i is given and not estimated. So the randomness in the sample estimate of (c23) comes from $\hat{\xi}_a$, $\hat{\xi}_b$, and , $\hat{\mu}$. LWA establish the joint asymptotic distribution of these three random variables. As outlined in Appendix B above, LWA (Theorem 2.1) establish that, under general regularity conditions, $\hat{\xi}_a$, $\hat{\xi}_b$, and , $\hat{\mu}$ are asymptotically joint normally distributed with (asymptotic) variance-covariance matrix $\Sigma = [\sigma_{ij}]$ where

$$\sigma_{11} = \frac{p_a(1-p_a)}{[f(\xi_a)]^2}, \qquad \sigma_{22} = \frac{p_b(1-p_b)}{[f(\xi_b)]^2}, \qquad \sigma_{33} = \sigma^2$$

$$\sigma_{12} = \frac{p_a(1-p_b)}{f(\xi_a)f(\xi_b)} = \sigma_{21} ,$$

and $\sigma_{13} = \frac{[\xi_a - \mu(1-p_a)]}{f(\xi_a)} = \sigma_{31}$

$$\sigma_{23} = \frac{[\xi_b - \mu(1-p_b)]}{f(\xi_b)} = \sigma_{32} \ .$$

Combining this set of results with Rao's linkage theorem then implies that $R\widehat{M}I_i = \hat{\mu}_i \ / \ \hat{\mu}$ is also asymptotically normally distributed with (asymptotic) variance

$$Asy.var\left(R\widehat{M}I_i\right) = g' \, \Sigma \, g \qquad \text{(c24)}$$

where, if $g = \left[\frac{\partial RMI_i}{\partial \xi_a}, \frac{\partial RMI_i}{\partial \xi_b}, \frac{\partial RMI_i}{\partial \mu}\right]' = [g_1, g_2, g_3]'$,

then $\frac{\partial RMI_i}{\partial \xi_a} = \left(\frac{1}{D_i}\right) \frac{\partial N_i}{\partial \xi_a}$

$$\frac{\partial RMI_i}{\partial \xi_b} = \left(\frac{1}{D_i}\right) \frac{\partial N_i}{\partial \xi_b}$$

and $\frac{\partial RMI_i}{\partial \mu} = \left(\frac{1}{D_i}\right) \frac{\partial N_i}{\partial \mu}$.

In the case of the lower-income group L,

$$g_1 = \frac{\partial RMI_L}{\partial \xi_a} = \left(\frac{1}{D_L}\right)\left[\left(\frac{1}{\mu}\right) \xi_a \bullet f(\xi_a)\right]$$

$$g_2 = \frac{\partial RMI_L}{\partial \xi_b} = 0$$

$$g_3 = \frac{\partial RMI_L}{\partial \mu} = \left(\frac{1}{D_L}\right)\left[-\left(\frac{1}{\mu^2}\right)\int_0^{\xi_a} y\, f(y)dy\right]$$

$$= \left(\frac{1}{D_L}\right)\left[-\left(\frac{1}{\mu}\right) \bullet N_L\right]$$

$$= \left(\frac{1}{D_L}\right)\left[-\left(\frac{D_L}{\mu}\right) \bullet \frac{N_L}{D_L}\right]$$

$$= -\left(\frac{1}{\mu}\right) \bullet RMI_L \ .$$

Therefore,

$$Asy.var\left(R\widehat{M}I_L\right) = g_1^2 \, \sigma_{11} + 2 g_1 g_3 \sigma_{13} + g_3^2 \, \sigma_{33}$$

$$\left(\frac{1-p_a}{p_a}\right)\left(\frac{\xi_a}{\mu}\right)^2 + \frac{RMI_L^2}{\mu^2} \bullet \sigma^2 + 2\left[\left(\frac{1-p_a}{p_a}\right)\left(\frac{\xi_a}{\mu}\right) \bullet RMI_L - \left(\frac{\xi_a}{\mu}\right)^2 \bullet \frac{RMI_L}{p_a}\right]. \qquad \text{(c25)}$$

In the case of the higher-income group *H*,

$$g_1 = \frac{\partial RMI_H}{\partial \xi_a} = 0$$

$$g_2 = \frac{\partial RMI_H}{\partial \xi_b} = \left(\frac{1}{D_H}\right)\left[-\left(\frac{1}{\mu}\right) \xi_b \bullet f(\xi_b)\right]$$

$$g_3 = \frac{\partial RMI_H}{\partial \mu} = \left(\frac{1}{D_H}\right)\left[-\left(\frac{1}{\mu^2}\right) \int_b^{\infty} y\, f(y) dy\right]$$

$$= \left(\frac{1}{D_H}\right)\left[-\left(\frac{1}{\mu}\right) \bullet N_H\right]$$

$$= \left(\frac{1}{D_H}\right)\left[-\left(\frac{D_H}{\mu}\right) \bullet \frac{N_H}{D_H}\right]$$

$$= -\left(\frac{1}{\mu}\right) \bullet RMI_H .$$

Therefore,

$$Asy.var\left(R\widehat{M}I_H\right) = g_2^2\, \sigma_{22} + 2 g_2 g_3 \sigma_{23} + g_3^2\, \sigma_{33}$$

$$= \left(\frac{p_b}{1-p_b}\right)\left(\frac{\xi_b}{\mu}\right)^2 + \frac{RMI_H^2}{\mu^2} \bullet \sigma^2 + 2\left[\left(\frac{\xi_b}{\mu}\right)^2 \bullet \frac{RMI_H}{(1-p_b)} - \left(\frac{\xi_b}{\mu}\right) RMI_H\right]. \quad \text{(c26)}$$

And in the case of the middle-income group *M*,

$$g_1 = \frac{\partial RMI_M}{\partial \xi_a} = \left(\frac{1}{D_M}\right)\left[-\left(\frac{1}{\mu}\right) \xi_a \bullet f(\xi_a)\right]$$

$$g_2 = \frac{\partial RMI_M}{\partial \xi_b} = \left(\frac{1}{D_M}\right)\left[\left(\frac{1}{\mu}\right) \xi_b \bullet f(\xi_b)\right]$$

$$g_3 = \frac{\partial RMI_M}{\partial \mu} = \left(\frac{1}{D_M}\right)\left[-\left(\frac{1}{\mu^2}\right) \int_{\xi_a}^{\xi_b} y\, f(y) dy\right]$$

$$= \left(\frac{1}{D_M}\right)\left[-\left(\frac{D_M}{\mu}\right) \bullet \frac{N_M}{D_M}\right]$$

$$= -\left(\frac{1}{\mu}\right) \bullet RMI_M .$$

Therefore,

$$Asy.var\left(R\widehat{M}I_M\right) = [g_1, g_2, g_3]\begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{bmatrix}\begin{bmatrix}g_1\\ g_2\\ g_3\end{bmatrix}$$

$$= g_1^2\,\sigma_{11} + g_2^2\,\sigma_{22} + g_3^2\,\sigma_{33} + 2g_1, g_2\,\sigma_{12} + 2g_1, g_3\,\sigma_{13} + 2g_2, g_3\,\sigma_{23}$$

$$= \left(\frac{1}{D_M^2}\right)\left[\left(\frac{\xi_a}{\mu}\right)^2 p_a(1-p_a)\right] + \left(\frac{1}{D_M^2}\right)\left[\left(\frac{\xi_b}{\mu}\right)^2 p_b(1-p_b)\right] + \left(\frac{RMI_M}{\mu}\right)^2 \bullet \sigma^2$$

$$- 2\left(\frac{1}{D_M^2}\right)\left[\left(\frac{\xi_a}{\mu}\right)\left(\frac{\xi_b}{\mu}\right) p_a(1-p_b)\right] \quad \text{(c27)}$$

$$+ 2\left(\frac{1}{D_M}\right)\left(\frac{\xi_a}{\mu}\right)\left(\frac{RMI_M}{\mu}\right)[\xi_a - \mu(1-p_a)]$$

$$- 2\left(\frac{1}{D_M}\right)\left(\frac{\xi_b}{\mu}\right)\left(\frac{RMI_M}{\mu}\right)[\xi_b - \mu(1-p_b)]\ ,$$

where $D_M = p_b - p_a$.

For all three cases, then,

$$S.E.\left(R\widehat{M}I_i\right) \equiv S.E.(\hat{\mu}_i\ /\ \hat{\mu}) = \left[\frac{Asy\hat{.}var(R\widehat{M}I_i)}{N}\right]^{1/2} \quad \text{(c28)}$$

for $i = L, M, H$.

The two approaches in this section result in two different representations (in eqs. (c22) and (c28)) for the standard errors of the relative-mean income ratios, but they are both valid, and the choice of use is up to the user.

C.4 Standard Errors of Income Shares

After the lengthy derivation of standard errors in the previous section, the derivation of standard errors for the income share estimates, $\widehat{IS}_i$, is straightforward. Note that, by identity,

$$\frac{\mu_i}{\mu} = \frac{IS_i}{PS_i}$$

where the population shares of the income groups in this appendix are given (non-random) and have been represented in the above derivations by the given D_i's. Therefore,

$$IS_i = D_i \bullet (\mu_i\ /\ \mu) = D_i \bullet RMI_i$$

for $i = L, M, H$. Similarly for their sample estimates:

$$\widehat{IS}_i = D_i \bullet R\widehat{M}I_i .$$

Thus it follows immediately that

$$Asy.var(\widehat{IS}_i) = D_i^2 \bullet Asy.var(R\widehat{M}I_i) \qquad \text{(c29)}$$

and hence

$$S.E.(\widehat{IS}_i) = D_i \bullet S.E.(R\widehat{M}I_i) \qquad \text{(c30)}$$

for *i* = *L, M, H.*

As a check on this result, it is still useful to work out the asymptotic variances of the income shares directly from first principles. This will also allow for an intuitive interpretation of the results.

From first principles,

$$IS_i = \int_{R_i} \left(\frac{1}{\mu}\right) y\, f(y) dy \qquad \text{for } i = L, M, H,$$

which is what we have been referring to as N_i in the previous section. Alternatively,

$$IS_i = D_i \bullet RMI_i = D_i \left[\frac{N_i(\xi_a, \xi_b, \mu)}{D_i}\right]$$

$$= N_i(\xi_a, \xi_b, \mu) .$$

So to work out asymptotic variances, one can apply the LWA results directly to $N_i(\xi_a, \xi_b, \mu)$. By LWA's Theorem 2.1, the $\widehat{IS}_i$ are asymptotically normally distributed each with an asymptotic variance given by

$$Asy.var(\widehat{IS}_i) = g' \Sigma g$$

with the same Σ matrix as before, but with the g vector's derivatives with respect to $IS_i =$ $N_i(\xi_a, \xi_b, \mu)$:

$$g = \left[\frac{\partial N_i}{\partial \xi_a}, \frac{\partial N_i}{\partial \xi_b}, \frac{\partial N_i}{\partial \mu}\right]' .$$

In the case of the lower-income group, *L*,

$$g_1 = \frac{\partial N_L}{\partial \xi_a} = \left(\frac{1}{\mu}\right) \xi_a \bullet f(\xi_a)$$

$$g_2 = \frac{\partial N_L}{\partial \xi_b} = 0$$

$$g_3 = \frac{\partial N_L}{\partial \mu} = -\left(\frac{1}{\mu^2}\right) \int_0^{\xi_a} y\, f(y) dy$$

$$= -\left(\frac{1}{\mu}\right) \bullet IS_L .$$

Therefore,

$$Asy.var\left(\widehat{IS}_L\right) = g_1^2\, \sigma_{11} + g_3^2\, \sigma_{33} + 2 g_1 g_3 \sigma_{13}$$

$$= p_a(1 - p_a)\left(\frac{\xi_a}{\mu}\right)^2 + \frac{IS_L^2}{\mu^2} \bullet \sigma^2 \qquad \text{(c31)}$$

$$- 2\left[\left(\frac{\xi_a}{\mu}\right) \bullet \frac{IS_L}{\mu}\right][\xi_a - \mu(1 - p_a)] .$$

In the case of the higher-income group, *H*,

$$g_1 = \frac{\partial N_H}{\partial \xi_a} = 0$$

$$g_2 = \frac{\partial N_H}{\partial \xi_b} = -\left(\frac{1}{\mu}\right) \xi_b \bullet f(\xi_b)$$

$$g_3 = \frac{\partial N_H}{\partial \mu} = -\left(\frac{1}{\mu^2}\right) \int_{\xi_b}^{\infty} y\, f(y) dy$$

$$= -\left(\frac{1}{\mu}\right) \bullet IS_H .$$

Therefore,

$$Asy.var\left(\widehat{IS}_H\right) = g_2^2\, \sigma_{22} + g_3^2\, \sigma_{33} + 2 g_2 g_3 \sigma_{23}$$

$$= p_b(1 - p_b)\left(\frac{\xi_b}{\mu}\right)^2 + \frac{IS_H^2}{\mu^2} \bullet \sigma^2 \qquad \text{(c32)}$$

$$+ 2\left[\left(\frac{\xi_b}{\mu}\right) \bullet \frac{IS_H}{\mu}\right][\xi_b - \mu(1 - p_b)] .$$

And in the case of the middle-income group *M*,

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

$$g_2 = \frac{\partial N_M}{\partial \xi_b} = \left(\frac{1}{\mu}\right) \xi_b \bullet f(\xi_b)$$

$$g_3 = \frac{\partial N_M}{\partial \mu} = -\left(\frac{1}{\mu}\right) \bullet IS_M .$$

Therefore,

$$Asy.var(\widehat{IS}_M) = [g_1, g_2, g_3] \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}$$

$$= g_1^2\, \sigma_{11} + g_2^2\, \sigma_{22} + g_3^2\, \sigma_{33} + 2g_1 g_2 \sigma_{12} + 2g_1 g_3 \sigma_{13} + 2g_2 g_3 \sigma_{23}$$

$$= p_a(1 - p_a)\left(\frac{\xi_a}{\mu}\right)^2 + p_b(1 - p_b)\left(\frac{\xi_b}{\mu}\right)^2 + \left(\frac{IS_M}{\mu}\right)^2 \bullet \sigma^2$$

$$-2\left[p_a(1 - p_b)\left(\frac{\xi_a}{\mu}\right)\left(\frac{\xi_b}{\mu}\right)\right] \qquad \text{(c33)}$$

$$+ 2\left(\frac{\xi_a}{\mu}\right)\left(\frac{IS_M}{\mu}\right) [\xi_a - \mu(1 - p_a)]$$

$$- 2\left(\frac{\xi_b}{\mu}\right)\left(\frac{IS_M}{\mu}\right) [\xi_b - \mu(1 - p_b)] .$$

The first three terms in (c33) capture the variability of $\hat{\xi}_a$, $\hat{\xi}_b$, and $\hat{\mu}$, respectively. The fourth term captures the covariation of $\hat{\xi}_a$ and $\hat{\xi}_b$. The fifth term picks up the covariation of $\hat{\xi}_a$ and $\hat{\mu}$, and the last term captures the covariation of $\hat{\xi}_b$ and $\hat{\mu}$.

Note also that the standard error formulas for the $\hat{\mu}_i$'s, the $R\widehat{M}I_i$'s, and the $\widehat{IS}_i$'s for percentile cut-offs are all distribution-free in that they do *not* require specifying an underlying density form for the sample data.

C.5 Standard Errors of Quantile Mean Income Gaps

Consider next the estimated quantile mean income gaps, $\hat{\mu}_i - \hat{\mu}_j$ for $i \neq j$ and *i, j = L, M, H.* Since this gap is the difference between quantile mean statistics, it follows that

$$Asy.var(\hat{\mu}_i - \hat{\mu}_j) = Asy.var(\hat{\mu}_i) - 2\,Asy.cov(\hat{\mu}_i, \hat{\mu}_j) + Asy.var(\hat{\mu}_j) \tag{c34}$$

where the variance terms are provided above in eqs. (c11)-(c13) and the covariance terms in eqs. (c14)-(c16). It then follows that

$$S.E.(\hat{\mu}_i - \hat{\mu}_j) = \left[\frac{Asy\hat{.}var(\hat{\mu}_i - \hat{\mu}_j)}{N}\right]^{1/2}. \tag{c35}$$

Similarly, if one calculates the gap in relative or proportional terms as

$$\hat{q} = \frac{\hat{\mu}_i - \hat{\mu}_j}{\mu_j} = \frac{\hat{\mu}_i}{\mu_j} - 1 \qquad \text{for} > j\,,$$

then it can be seen that

$$Asy.var(\hat{q}) = \left(\frac{1}{\mu_j}\right)^2 \bullet Asy.var(\hat{\mu}_i) + \left(\frac{\mu_i}{\mu_j^2}\right)^2 \bullet Asy.var(\hat{\mu}_j)$$

$$- 2\left(\frac{1}{\mu_j}\right)\left(\frac{\mu_i}{\mu_j^2}\right) \bullet Asy.cov(\hat{\mu}_i, \hat{\mu}_j)\,, \tag{c36}$$

so that again

$$S.E.(\hat{q}) = \left[\frac{Asy\hat{.}var(\hat{q})}{N}\right]^{1/2}. \tag{c37}$$

<u>C.6 Standard Errors of Relative-Mean Quantiles</u>

A final statistic of potential interest in a disaggregative analysis is the ratio of a quantile income cut-off level to overall mean of a distribution $(\hat{\xi}_i / \hat{\mu})$, or relative-mean quantiles. Here we again follow the work of Lin, Wu and Ahmad (1980). Consider two percentile cut-off levels corresponding to the proportions $p_a < p_b$; call these ξ_a and ξ_b. Then, under a well-behaved continuous density $(\bullet)$, the authors establish that the set of sample relative-mean quantiles

$\hat{\xi}_a$ / $\hat{\mu}$ and $\hat{\xi}_b$ / $\hat{\mu}$ are asymptotically joint normally distributed with (asymptotic) variances and covariances:

$$Asy.var(\hat{\xi}_a / \hat{\mu}) = \left[\left(\frac{\xi_a}{\mu}\right)^2 \sigma^2 - 2\left(\frac{\xi_a}{\mu}\right)\sigma_{13} + \sigma_{11}\right] / \mu^2 \quad \text{(c38)}$$

$$Asy.var(\hat{\xi}_b / \hat{\mu}) = \left[\left(\frac{\xi_b}{\mu}\right)^2 \sigma^2 - 2\left(\frac{\xi_b}{\mu}\right)\sigma_{23} + \sigma_{22}\right] / \mu^2 \quad \text{(c39)}$$

and

$$Asy.cov(\hat{\xi}_a / \hat{\mu}, \hat{\xi}_b / \hat{\mu}) = \left[\left(\frac{\xi_a \bullet \xi_b}{\mu^2}\right)\sigma^2 - \left(\frac{\xi_a}{\mu}\right)\sigma_{23} - \left(\frac{\xi_b}{\mu}\right)\sigma_{13} + \sigma_{12}\right] / \mu^2 \quad \text{(c40)}$$

where $\sigma_{11} = \frac{p_a(1-p_a)}{f(\xi_a)^2}$, $\sigma_{22} = \frac{p_b(1-p_b)}{f(\xi_b)^2}$,

$\sigma_{12} = \frac{p_a(1-p_b)}{f(\xi_a)\bullet f(\xi_b)}$

and $\sigma_{13} = \frac{\xi_a - \mu(1-p_a)}{f(\xi_a)}$ $\sigma_{23} = \frac{\xi_b - \mu(1-p_b)}{f(\xi_b)}$.

Therefore, it follows that

$$S.E.(\hat{\xi}_i / \hat{\mu}) = \left[\frac{Asy\hat{.}var(\hat{\xi}_i / \hat{\mu})}{N}\right]^{1/2} \quad \text{(c41)}$$

for *i* = *a, b*, where all unknowns are replaced by their sample estimates.

Note that these standard error formulas are definitely not distribution-free as they depend upon the density ordinates in the denominators of the σ_{ij} terms.

Appendix D

Appendix Table D1
Summary Statistics on Annual Earnings for Census Estimation Samples and Median-Based Cut-Offs for Full-Time Workers
Selective Years 1970-2005
(real 2015 dollars)

	1970	1980	1990	2000	2005
Males					
No. obs.	28,405	68,614	122,859	121,923	124,231
Mean earnings	55,552	65,238	64,612	65,366	74,474
Median earnings	49,552	59,490	58,468	57,258	58,310
MC earnings range	24,776-74,329	29,745-89,236	29,234-87,702	28,629-85,888	29,155-87,465
Mean MC earnings	47,971	58,713	57,166	55,446	56,151
Lower-earnings cut-off	24,776	29,745	29,234	28,629	29,155
Higher earnings cut-off	99,104	118,981	116,937	114,516	116,620
Females					
No. obs.	8,608	30,653	78,693	87,871	94,693
Mean earnings	32,932	42,065	44,076	47,279	52,076
Median earnings	30,735	39,068	40,603	42,710	44,030
MC earnings range	15,368-46,102	19,534-58,600	20,301-60,904	21,355-64,065	22,015-66,045
Mean MC earnings	29,846	38,278	39,833	41,451	42,528
Lower-earnings cut-off	15,368	19,534	20,301	21,355	22,015
Higher earnings cut-off	61,469	78,135	81,206	85,421	88,060

Note: Based on Census public use microdata files.
Inflation adjustment based on CPI.

Appendix Table D2
Summary Statistics on Weekly Earnings for LFS Estimation Samples and Median-Based Cut-Offs for Full-Time Workers
Selective Years 2000-2015
(real 2015 dollars)

	2000	2005	2010	2015
Males				
No. obs.	19,476	19,047	19,268	36,678
Mean earnings	1106.5	1110.5	1185.1	1215.5
Median earnings	1025.3	1021.5	1075.8	1105.4
MC earnings range	512.7-1538.0	510.7-1532.2	537.9-1613.6	552.7-1657.9
Mean MC earnings	990.4	987.2	1031.2	1051.2
Lower-earnings cut-off	512.7	510.7	537.9	552.7
Higher earnings cut-off	2050.7	2043.0	2151.5	2210.6
Females				
No. obs.	14,979	15,842	17,105	32,052
Mean earnings	846.5	880.6	959.9	998.0
Median earnings	769.0	798.1	863.2	881.8
MC earnings range	384.6-1153.6	399.0-1197.1	431.6-1294.9	441.0-1322.8
Mean MC earnings	739.1	761.4	814.5	827.8
Lower earnings cut-off	384.6	399.0	431.6	441.0
Higher earnings cut-off	1538.0	1596.1	1726.3	1763.7

Note: Based on May Labour Force Surveys.

Appendix Table D3
Summary Statistics on Percentile Annual Earnings Samples from Census Data for Full-Time Workers
Selective Years 1970-2005
(real 2015 dollars)

	1970	1980	1990	2000	2005
Males					
20th percentile	33,934	41,065	38,427	35,084	34,510
Mean MC earnings	49,931	60,634	59,445	58,706	60,170
80th percentile	68,996	83,922	84,453	86,756	92,820
Females					
20th percentile	21,326	27,477	25,986	26,694	26,180
Mean MC earnings	30,919	39,867	41,529	43,918	45,719
80th percentile	43,279	56,141	60,091	66,735	70,210

Appendix Table D4
Summary Statistics on Percentile Weekly Earnings Samples from LFS Data for Full-Time Workers
Selective Years 2000-2015
(real 2015 dollars)

	2000	**2005**	**2010**	**2015**
Males				
20^{th} percentile	694.5	686.0	719.3	719.2
Mean MC earnings	1044.9	1037.6	1104.3	1129.6
80^{th} percentile	1438.2	1466.8	1592.9	1632.7
Females				
20^{th} percentile	514.2	533.3	572.2	599.3
Mean MC earnings	794.0	814.6	890.9	922.9
80^{th} percentile	1145.3	1187.8	1299.6	1370.7