Queen's Economics Department Working Paper No. 1477

A Nifty Fix for Published Distribution Statistics: Simplified Distribution-Free Statistical Inference

Charles Beach

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

9-2021

A Nifty Fix for Published Distribution Statistics: Simplified Distribution-Free Statistical Inference

September 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 and Sharon Sullivan. All errors and limitations of this paper, though, are solely the author's.

Abstract

This paper applies the tool box measures of disaggregative income inequality characterization and the statistical methodology of Beach (2021) to percentile-based distribution statistics such as quintile income shares and decile means typically published by official statistical agencies. It derives standard error formulas for those measures which are distribution-free and easy to implement. The approach is illustrated with Canadian Labour Force Survey data over 1997-2015. It is found that widely shared real earnings gains were experienced over this period, but that the gains were very unevenly shared with middle-class workers losing out relatively and top earners having highly statistically significant earnings gains.

1. Introduction

At a time of dramatic changes to the economy, labour markets and social policy in response to the wide-spread COVID-19 epidemic, the need is ever greater for disaggregative measures of income distributional changes and disaggregative tools to characterize separate regions of an income distribution. Such tools help in describing what has been happening in different regions of the income distribution (e.g., middle-class workers vs very high earners and income recipients) and allow a focus on what has been happening among different groups across the lower end of the distribution (e.g., low-wage workers vs retirees and those on government support programs). These tools serve as a prelude to designing appropriate and targeted policies to address concerns, and help in testing their effectiveness. Alternative explanations or hypotheses on the principal drivers of distributional changes and the channels through which they operate also need detailed measures to characterize different regions of the distribution for various groups and thence be able to test among these alternative hypotheses. These needs point to (i) the usefulness of a set of disaggregative measures to characterize detailed distributional changes or differences between groups, and (ii) the importance of being able to perform formal statistical inference with these measures.

In an earlier paper (Beach, 2021), the author forwarded a tool box of just such disaggregative measures for characterizing detailed distributional changes and differences, and developed a "quantile function approach" that allowed for formal statistical inference with each of the tool box measures. The present paper makes use of this tool box set of measures and applies the quantile function approach to percentile-based disaggregative inequality measures (such as decile or quintile income shares) typically published by official government statistical agencies such as Statistics Canada or the United States Bureau of the Census.[1] It thus provides a relatively simple representation of the (asymptotic) distribution of these estimated measures and their asymptotic variances (and co-variances), and thence standard error formulas that can be readily applied to such published official statistics. This can help the empirical user of these standard statistics to see how meaningful or reliable any observed differences or changes indeed are. The paper thus extends the quantile function approach to percentile-based tool box measures for characterizing an income distribution, and provides an easy-to-use representation of (asymptotic) variances and standard errors for these measures. In deriving the latter, it shows that the quantile function approach leads to distribution-free standard error formulas for these percentile-based measures, so applied users do not need to engage in burdensome kernel estimation techniques or make restrictive assumptions about the specific functional form of the underlying income distribution.

The paper proceeds as follows. The next section presents the derivation of the quantile-function approach standard errors formulas that are used. Section 3 illustrates some empirical applications of these formulas using Canadian earnings data for 1997-2015. Section 4 then concludes with an overview of the paper and a review of its principal findings and recommendations.

[1] See, for example the quintile and decile income share data series from Statistics Canada's CANSIM Table 206-0031.

2. Derivation of Standard Error Formulas for Percentile Statistics

2.1 Percentile Statistics and Tool Box Measures

Percentile statistics are those that are expressed in terms of given percentage groups of the ranked or ordered observations in a sample. In the case of income distribution statistics, the data observations in a sample are ordered by income from the lowest income observation to the highest income observation. The ordered observations are then divided into non-overlapping income groups, say in terms of deciles or quintiles (or generically referred to as quantiles). So the first decile group consists of these observations with the 10 percent lowest income levels, the second decile group consists of the next 10 percent lowest income recipients, and so on up to the top or tenth decile group which includes those 10 percent of income recipients with the highest income levels in the sample. Similarly, quintiles consist of five income groups ordered from the first or lowest-income or bottom 20 percent of recipients up to the fifth or highest-income or top 20 percent of income recipients. The standard Lorenz curve, for example, is based around such percentile groups. The key feature of such percentile statistics is that the relative sizes of the percentile groups are *given* percentages of the sample or distribution.

The disaggregative tool box measures for characterizing changes and differences in income distributions set out in Beach (2021) include:

- income shares
- quantile means and percentile income cut-offs
- quantile income gaps and differentials
- relative mean incomes.

Income shares are the proportions of total income in the distribution being received by members within a particular income group (e.g., the top decile of the distribution or by the Middle Class often defined as the middle 60 percent of income recipients). Percentile cut-off statistics are the income levels that separate one percentile income group from an adjacent one. So the first decile cut-off is that income level that separates income recipients in the lowest and the second decile income groups. Quantile means (sometimes referred to as conditional means) are the average or mean incomes of the income recipients within a given quantile group. So the mean middle-class income is the average of all incomes belonging to the middle-class income group (say, the middle 60 percent of recipients). Income gaps are the differences between the quantile mean incomes of two specified income groups (e.g., between mean middle-class income and the top decile mean income level). Income differentials are the ratios or percentage differences between quantile mean incomes. Relative mean incomes are the ratios between quantile-specific means and the overall mean income of the distribution. Income gaps and differentials can also be calculated between different distributions, such as between male and female earners in the labour market. These are illustrated in the empirical tables examined in Section 3 below.

2.2 The Quantile Function Approach to Estimating Standard Errors of Tool Box Measures

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

Consider first some formal concepts and notation. Suppose the distribution of income *Y*, is divided into *K* ordered income groups, so that *K = 10* in the case of deciles and *K = 5* for quintiles. Let the dividing proportions of recipients be $p_1 < p_2 < \cdots < p_{K-1}$ (with $p_0 = 0$ and $p_K = 1.0$).[2] Then in terms of the underlying (population) density of income recipients, the mean income of the i'th quantile is given by

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

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

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

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

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

[2] 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.

asymptotic distribution of the sample cut-offs $\hat{\xi}_i$'s has long been well established, this theorem provides the basis of the quantile function approach (or QFA) used in Beach (2021) and the present paper. The basic idea is to express the various percentile tool box measures in terms of integral functions of the income cut-offs (the ξ_i's) and then invoke Rao's linkage theorem to establish asymptotic normality and expressions for the sample measures' asymptotic variances. Standard errors, then, are simply obtained from these estimated (asymptotic) variances rescaled by the size of the estimation sample:

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

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

Now, in general one would expect the (asymptotic) variances to depend on the specific form of the underlying income distribution's density $f(\bullet)$. Certainly the (asymptotic) variance-covariance structure of the $\hat{\xi}_i$'s does. But – as will be shown in the next several subsections – perhaps surprisingly, the resulting (asymptotic) variances and standard errors of the percentile-based tool box measures are a special case that do *not* depend upon the specific function form of $f(\bullet)$. In this sense, they are said to be distribution-free, and hence very straightforward to calculate. As a result, the "nifty fix" cited in this paper's title refers to the simple information that can be added to published official distribution statistics to usefully indicate the reliability of the (sample) survey estimates.

2.3 Application of QFA to Conditional Means

The starting point is to establish the asymptotic distribution and its variance-covariance structure for the full set of sample quantile income cut-off levels. Suppose that the income distribution is divided into *K* ordered income groups corresponding to the cumulative proportions

$0 < p_1 < p_2 < \cdots < p_K = 1$ and the quantile cut-offs $\xi_1, \xi_2, \ldots \xi_{K-1}$. Let $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \ldots \hat{\xi}_{K-1})'$ be a vector of *K-1* sample quantile cut-offs[3] from a random sample of size N drawn from a continuous population density $f(\bullet)$ such that the $\hat{\xi}_i$'s are uniquely defined and $f_i \equiv f(\xi_i) > 0$ for all *i = 1, ..., K-1*. Then it can be proved (see, for example, Wilks (1962), p. 273, or Kendall and Stuart (1969, pp. 237-239)) that the vector $\sqrt{N}\,(\hat{\xi} - \xi)$ converges in distribution to a *(K-1)*-variate normal distribution with mean zero and variance-covariance matrix **Λ** where

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

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

Then applying a multivariate version of Rao's linkage theorem (Rao, 1965, p. 388), consider the full set of sample quantile means $\widehat{m} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K)'$ corresponding to the vector of population quantile means $m = (\mu_1, \mu_2, \ldots, \mu_K)'$ where μ_i is defined in eq. (2). In the case of deciles, *K = 10* and, $D_i = 0.10$, and in the case of quintiles, *K = 5* and $D_i = 0.20$. Then according to Rao's theorem for continuous differentiable functions, the vector $\widehat{m}$ is asymptotically joint normally distributed in that $\sqrt{N}(\widehat{m} - m)$ converges in distribution to a joint normal with *KxK* (asymptotic) variance-covariance matrix *V* where

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

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

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

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

$$= \left[\frac{\partial \mu_i}{\partial \xi_j}\right] \qquad \text{with } i = 1, \ldots, K \text{ rows and } j = 1, \ldots, K\text{-}1 \text{ columns.} \tag{4b}$$

For convenience, rewrite eq. (2) as

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

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

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

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

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

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

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

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

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

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

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

$$G = \left[\begin{array}{ccccccc} 10\,\xi_1 \bullet f(\xi_1) & 0 & 0 & 0 & \cdots & 0 & 0 \\ -10\,\xi_1 \bullet f(\xi_1) & 10\,\xi_2 \bullet f(\xi_2) & 0 & 0 & \cdots & 0 & 0 \\ 0 & -10\,\xi_2 \bullet f(\xi_2) & 10\,\xi_3 \bullet f(\xi_3) & 0 & \cdots & \vdots & \vdots \\ \vdots & 0 & -10\,\xi_3 \bullet f(\xi_3) & 10\,\xi_4 \bullet f(\xi_4) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & -10\,\xi_8 \bullet f(\xi_8) & 10\,\xi_9 \bullet f(\xi_9) \\ 0 & 0 & 0 & 0 & \cdots & 0 & -10\,\xi_9 \bullet f(\xi_9) \end{array}\right] . \tag{5}$$

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

$$Asy.var(\hat{\mu}_1) = G(row\ 1) \bullet \Lambda \bullet\ G(row\ 1)'$$

$$= (10)^2\, \xi_1^2\ \bullet f(\xi_1)^2\ \bullet\ \left[\frac{p_1(1-p_1)}{f(\xi_1)^2}\right]$$

$$= (10)^2\, p_1(1-p_1)\, \xi_1^2\ . \tag{6a}$$

Similarly,

$$Asy.var(\hat{\mu}_{10}) = G(row\ 10) \bullet \Lambda \bullet\ G(row\ 10)'$$

$$= (10)^2\, \xi_9^2\ \bullet f(\xi_9)^2\ \bullet\ \left[\frac{p_9(1-p_9)}{f(\xi_9)^2}\right]$$

$$= (10)^2\, p_9(1-p_9)\, \xi_9^2\ . \tag{6b}$$

And for $i = 2, \ldots, 9$,

$$Asy.var(\hat{\mu}_i) = G(row\ i) \bullet \Lambda \bullet\ G(row\ i)'$$

$$= (10)^2\, [p_{i-1}(1-p_{i-1})\xi_{i-1}^2 + p_i(1-p_i)\, \xi_i^2$$

$$- 2\, p_{i-1}(1-p_i)\xi_{i-1}\, \xi_i]\ . \tag{6c}$$

More generally, then,

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

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

and for *i = 2, …, K-1.*

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

$$- 2\left(\frac{1}{D_{i-1}}\right)\left(\frac{1}{D_i}\right) p_{i-1}(1-p_i)\xi_{i-1}\ \xi_1 \ . \qquad (7)$$

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

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

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

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

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

for *i = 2, …, K-1.*

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

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

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

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

Or more generally,

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

for all $i = 1, \ldots, K$.

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

2.4 Quantile Mean Income Gaps and Differentials

One question that practitioners may be interested in is whether the income gap between, say, middle and top incomes has changed significantly over time. To address this question requires information not just on variances, but also on covariances between estimates of middle and top incomes. Conveniently, the general results in eqs. (4) and (5) above allow one to provide an answer in the case of quantile means.

The way to calculate (asymptotic) covariances from eqs. (4) and (5) is the same as for the variances except that, since the covariances are the off-diagonal elements in eq. (4), the calculations involve using *different* rows of matrix *G*. Again, let us illustrate this in the case of decile income groups (i.e., $D_i = 0.10$ for all $i = 1, \ldots, 10$). Consider, for example,

$$\begin{aligned} Asy.cov(\hat{\mu}_3, \hat{\mu}_5) &= G(row\ 3) \bullet \Lambda \bullet\ G(row\ 5)' \\ &= \big(-10\xi_2 \bullet f(\xi_2)\big)\big(-10\xi_4 \bullet f(\xi_4)\big)\lambda_{24} + \big(-10\xi_2 \bullet f(\xi_2)\big)\big(10\xi_5 \bullet f(\xi_5)\big)\lambda_{25} \\ &\quad + \big(10\xi_3 \bullet f(\xi_3)\big)\big(-10\xi_4 \bullet f(\xi_4)\big)\lambda_{34} + \big(10\xi_3 \bullet f(\xi_3)\big)\big(10\xi_5 \bullet f(\xi_5)\big)\lambda_{35} \\ &= (10)^2[p_2(1-p_4)\xi_2\xi_4 - p_2(1-p_5)\xi_2\xi_5 - p_3(1-p_4)\xi_3\xi_4 \\ &\quad + p_3(1-p_5)\xi_3\xi_5]\ . \end{aligned}$$

More generally:

For $1 < i < j < 10$:

$$Asy.cov(\hat{\mu}_i, \hat{\mu}_j) = (10)^2[p_{i-1}(1-p_{j-1})\xi_{i-1}\xi_{j-1} - p_{i-1}(1-p_j)\xi_{i-1}\xi_j - p_i(1-p_{j-1})\xi_i\xi_{j-1} + p_i(1-p_j)\xi_i\xi_j] \quad (10a)$$

For $1 = i < j < 10$:

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_j) = (10)^2[-p_1(1-p_{j-1})\xi_1\xi_{j-1} + p_1(1-p_j)\xi_1\xi_j] \quad (10b)$$

For $1 < i < j = 10$:

$$Asy.cov(\hat{\mu}_i, \hat{\mu}_{10}) = (10)^2[p_{i-1}(1-p_9)\xi_{i-1}\xi_9 - p_i(1-p_9)\xi_i\xi_9] \quad (10c)$$

For $1 = i < j = 10$:

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_{10}) = (10)^2[-p_1(1-p_9)\xi_1\xi_9] \, . \quad (10d)$$

So for the $(\hat{\mu}_{10} - \hat{\mu}_5)$ mean income gap, the (asymptotic) covariance is

$$Asy.cov(\hat{\mu}_5, \hat{\mu}_{10}) = (10)^2[p_4(1-p_9)\xi_4\xi_9 - p_5(1-p_9)\xi_5\xi_9] \, . \quad (11a)$$

The $\hat{\mu}_5 - \hat{\mu}_1$ mean income gap covariance is

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_5) = (10)^2[-p_1(1-p_4)\xi_1\xi_4 + p_1(1-p_5)\xi_1\xi_5] \, . \quad (11b)$$

And the (asymptotic) covariance for the $\hat{\mu}_{10} - \hat{\mu}_1$ mean income gap is given in eq. (10d) above.

Since the income gap is a linear function of random variables, it follows that, for example,

$$Asy.var(\hat{\mu}_{10} - \hat{\mu}_5) = Asy.var(\hat{\mu}_5) + Asy.var(\hat{\mu}_{10}) - 2\, Asy.cov(\hat{\mu}_5, \hat{\mu}_{10}) \, . \quad (12)$$

So then

$$S.E.(\hat{\mu}_{10} - \hat{\mu}_5) = \left[\frac{As\hat{y.v}ar(\hat{\mu}_{10} - \hat{\mu}_5)}{N}\right]^{1/2} \quad (13)$$

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

By a quantile mean income differential is meant the proportional difference between two quantile means; for example,

$$\hat{q} = (\hat{\mu}_{10} - \hat{\mu}_5) \; / \; \hat{\mu}_5 = \left(\frac{\hat{\mu}_{10}}{\hat{\mu}_5}\right) - 1 \; .$$

While this relationship is certainly not linear, it is still continuous and differentiable in its arguments, so Rao's linkage theorem again applies. We have already established the joint asymptotic normality of $\hat{\mu}_5$ and $\hat{\mu}_{10}$ and worked out their (asymptotic) covariance and variances. So, by Rao's theorem, $\hat{q}$ is also asymptotically normally distributed with (asymptotic) variance given by

$$Asy.var(\hat{q}) = Q \; V \; Q'$$

where here

$$V = \begin{bmatrix} Asy.var(\hat{\mu}_5) & Asy.cov(\hat{\mu}_5, \hat{\mu}_{10}) \\ Asy.cov(\hat{\mu}_5, \hat{\mu}_{10}) & Asy.var(\hat{\mu}_{10}) \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{bmatrix}$$

and $\quad Q = \left[\frac{\partial q}{\partial \mu_5}, \frac{\partial q}{\partial \mu_{10}}\right]$

with $\frac{\partial q}{\partial \mu_5} = \frac{-\mu_{10}}{\mu_5^2}$ and $\frac{\partial q}{\partial \mu_{10}} = \frac{1}{\mu_5}$.

Therefore,

$$\begin{aligned} Asy.var(\hat{q}) &= \left(\frac{-\mu_{10}}{\mu_5^2}\right)^2 \bullet \lambda_{11} + \left(\frac{1}{\mu_5}\right)^2 \bullet \lambda_{22} + 2\left(\frac{-\mu_{10}}{\mu_5^2}\right)\left(\frac{1}{\mu_5}\right) \bullet \lambda_{12} \\ &= \left(\frac{-\mu_{10}}{\mu_5^2}\right)^2 \bullet Asy.var(\hat{\mu}_5) + \left(\frac{1}{\mu_5}\right)^2 \bullet Asy.var(\hat{\mu}_{10}) \qquad (14) \\ &\qquad + 2\left(\frac{-\mu_{10}}{\mu_5^2}\right)\left(\frac{1}{\mu_5}\right) \bullet Asy.cov(\hat{\mu}_5, \hat{\mu}_{10}) \; , \end{aligned}$$

and consequently,

$$S.E.(\hat{q}) = \left[\frac{Asy\hat{.}var(\hat{q})}{N}\right]^{1/2} , \qquad (15)$$

where, as usual, all unknowns are replaced by their sample estimates.

2.5 Application of QFA to Relative Mean Incomes

By relative mean income is meant the ratio of a quantile mean income level divided by the overall mean income level of the distribution of incomes;

i.e.: $RMI_i = \frac{\mu_i}{\mu}$ for $i = 1, \ldots, K$, (16a)

with sample estimate

$$R\widehat{M}I_i = \frac{\hat{\mu}_i}{\hat{\mu}}. \tag{16b}$$

There are two alternative approaches that can be used to work out the asymptotic distribution of $R\widehat{M}I_i$.

1) Adding-Up Approach:

This approach recognizes that the overall mean of the income distribution is the average of the full set of quantile group means:

$$\mu = \sum_{i=1}^{K} D_i \mu_i \quad \text{and} \quad \hat{\mu} = \sum_{i=1}^{K} D_i \hat{\mu}_i \ .$$

Since the ratio $\mu_i \ / \ \mu$ is continuous and differentiable in its arguments, one can again apply the Rao linkage theorem to establish the (asymptotic) normality of $R\widehat{M}I_i$ with

$$Asy.var\left(R\widehat{M}I_i\right) = Q' \, W \, Q$$

where $Q = [q_1, q_2]'$

with $q_1 = \frac{\partial RMI_i}{\partial \mu_i} = \frac{1}{\mu}$ and $q_2 = \frac{\partial RMI_i}{\partial \mu} = \frac{-\mu_i}{\mu^2}$

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

Therefore,

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

$$+ 2\left(\frac{1}{\mu}\right)\left(\frac{-\mu_i}{\mu^2}\right) \bullet Asy.cov(\hat{\mu}_i, \hat{\mu}) \ . \tag{17}$$

Now $Asy.var(\hat{\mu}) = \sigma^2$ and $Asy.var(\hat{\mu}_i)$ has been derived above in eqs. (7) and (8). The $Asy.cov(\hat{\mu}_i, \hat{\mu})$ remains to be determined. But this is straightforward using the adding-up constraint. We illustrate in the case of decile income groups where $D_i = 0.10$ for all *i = 1, …, 10.*

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

Substituting in the expressions for $\hat{\mu}$ and μ leads to

$$Cov(\hat{\mu}_i, \hat{\mu}) = (.10)\left[Var(\hat{\mu}_i) + \textstyle\sum_{j \neq i} Cov(\hat{\mu}_i, \hat{\mu}_j)\right] .$$

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

$$Asy.cov(\hat{\mu}_i, \hat{\mu}) = (.10)\left[Asy.var(\hat{\mu}_i) + \textstyle\sum_{j \neq i} Asy.cov(\hat{\mu}_i, \hat{\mu}_j)\right] . \tag{18}$$

With all the (asymptotic) variances and covariances derived in the previous sections, expressions for all the terms in (18) are known, and hence (18) can be substituted into the third term of eq. (17) above.

2) <u>Joint Distribution Approach:</u>

This approach explicitly incorporates the joint distribution of the $\hat{\mu}_i$'s and $\hat{\mu}$. To do so, it makes use of a useful paper by Lin, Wu and Ahmad (1980) (henceforth LWA).

If one goes back to basics,

$$RMI_i \equiv \frac{\mu_i}{\mu} = \int_{R_i} \left(\frac{1}{\mu}\right) yf(y)dy \ / \int_{R_i} f(y)dy$$

$$= \left(\frac{1}{D_i}\right) \int_{R_i} \left(\frac{1}{\mu}\right) yf(y)dy = \left(\frac{1}{D_i}\right) N_i(\xi_i, \xi_{i-1}, \mu) \tag{19}$$

where R_i indicates the relevant range of integration for the quantile income group i, and N_i indicates that the integration expression is explicitly a function of the triplet of parameters $\xi_i, \xi_{i-1},$ and μ for *i = 2, ..., K-1* (or a doublet of parameters in the cases of *i=1* and *K*). Again, lets focus on the case of deciles, so that *K=10* and $\left(\frac{1}{D_i}\right) = 10$ as well.

LWA establish that, under general regularity conditions, $\hat{\xi}_i, \hat{\xi}_{i-1},$ and $\hat{\mu}$ are asymptotically joint normally distributed with (asymptotic) variance-covariance matrix

$$\Sigma = \left[\sigma_{ij}\right]$$

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

$$\sigma_{12} = \frac{p_{i-1}(1-p_i)}{f(\xi_{i-1})f(\xi_i)} = \sigma_{21} \tag{20}$$

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

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

Now combine this set of LWA results with Rao's linkage theorem. Together these imply that $R\widehat{M}I_i$ is also asymptotically normally distributed with (asymptotic) variance

$$Asy.var\left(R\widehat{M}I_i\right) = G' \, \Sigma \, G \tag{21}$$

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

Then $\frac{\partial RMI_i}{\partial \xi_{i-1}} = 10 \bullet \frac{\partial N_i}{\partial \xi_{i-1}}$

$$\frac{\partial RMI_i}{\partial \xi_i} = 10 \bullet \frac{\partial N_i}{\partial \xi_i}$$

and $\frac{\partial RMI_i}{\partial \mu} = 10 \bullet \frac{\partial N_i}{\partial \mu}$.

In the case of *i = 2, ..., 9*:

$$g_1 = 10 \bullet \left(\frac{-1}{\mu}\right) \xi_{i-1} f(\xi_{i-1})$$

$$g_2 = 10 \bullet \left(\frac{1}{\mu}\right) \xi_i f(\xi_i)$$

and $$g_3 = 10 \bullet \left(\frac{-1}{\mu}\right) \bullet N_i = 10 \left[-\left(\frac{1}{\mu}\right)\left(\frac{RMI_i}{10}\right)\right]$$

$$= -\left(\frac{RMI_i}{\mu}\right).$$

Therefore,

$$Asy.var\left(R\widehat{M}I_i\right) = G' \Sigma G$$

$$= g_1^2 \sigma_{11} + g_2^2 \sigma_{22} + g_3^2 \sigma_{33} + 2g_1 g_2 \sigma_{12} + 2g_1 g_3 \sigma_{13} + 2g_2 g_3 \sigma_{23}$$

$$= \left[10\left(\frac{-1}{\mu}\right) \xi_{i-1} \bullet f(\xi_{i-1})\right]^2 \bullet \left[\frac{p_{i-1}(1-p_{i-1})}{[f(\xi_{i-1})]^2}\right]$$

$$+ \left[10\left(\frac{1}{\mu}\right) \xi_i \bullet f(\xi_i)\right]^2 \bullet \left[\frac{p_i(1-p_i)}{[f(\xi_i)]^2}\right] + \left[-\left(\frac{RMI_i}{\mu}\right)\right]^2 \bullet \sigma^2$$

$$+2\left[10\left(\frac{-1}{\mu}\right) \xi_{i-1} \bullet f(\xi_{i-1})\right]\left[10\left(\frac{1}{\mu}\right) \xi_i \bullet f(\xi_i)\right] \bullet \left[\frac{p_{i-1}(1-p_i)}{f(\xi_{i-1}) \bullet f(\xi_i)}\right]$$

$$+2\left[10\left(\frac{-1}{\mu}\right) \xi_{i-1} \bullet f(\xi_{i-1})\right]\left[-\left(\frac{RMI_i}{\mu}\right)\right] \bullet \left[\frac{\xi_{i-1} - \mu(1-p_{i-1})}{f(\xi_{i-1})}\right]$$

$$+2\left[10\left(\frac{1}{\mu}\right) \xi_i \bullet f(\xi_i)\right]\left[-\left(\frac{RMI_i}{\mu}\right)\right] \bullet \left[\frac{\xi_i - \mu(1-p_i)}{f(\xi_i)}\right]$$

$$= 10^2\left[\left(\frac{\xi_{i-1}}{\mu}\right)^2 p_{i-1}(1-p_{i-1})\right] + 10^2\left[\left(\frac{\xi_i}{\mu}\right)^2 p_i(1-p_i)\right] + \left(\frac{RMI_i}{\mu}\right)^2 \sigma^2$$

$$-2(10)^2\left[\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{\xi_i}{\mu}\right) p_{i-1}(1-p_i)\right]$$

$$+2(10)\left[\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{RMI_i}{\mu}\right)[\xi_{i-1} - \mu(1-p_{i-1})]\right]$$

$$-2(10)\left[\left(\frac{\xi_i}{\mu}\right)\left(\frac{RMI_i}{\mu}\right)[\xi_i - \mu(1-p_i)]\right]. \qquad (22)$$

In the case of $i = 1$:

$$g_1 = 0$$

$$g_2 = 10 \bullet \left(\frac{1}{\mu}\right) \xi_i \bullet f(\xi_i)$$

$$g_3 = -\left(\frac{RMI_i}{\mu}\right).$$

So

$$Asy.var\left(R\widehat{M}I_1\right) = g_2^2\,\sigma_{22} + g_3^2\,\sigma_{33} + 2g_2g_3\sigma_{23}$$

$$= \left[10\left(\frac{1}{\mu}\right)\,\xi_1 \bullet f(\xi_1)\right]^2 \bullet \left[\frac{p_1(1-p_1)}{[f(\xi_1)]^2}\right] + \left[-\left(\frac{RMI_1}{\mu}\right)\right]^2 \bullet \sigma^2$$

$$+ 2\left[10\left(\frac{1}{\mu}\right)\,\xi_1 \bullet f(\xi_1)\right]\left[-\left(\frac{RMI_1}{\mu}\right)\right] \bullet \left[\frac{[\xi_1 - \mu(1-p_1)]}{f(\xi_1)}\right]$$

$$= 10^2\left[\left(\frac{\xi_1}{\mu}\right)^2 p_1(1-p_1)\right] + \left(\frac{RMI_1}{\mu}\right)^2 \bullet \sigma^2$$

$$-2(10)\left[\left(\frac{\xi_1}{\mu}\right)\left(\frac{RMI_1}{\mu}\right)[\xi_1 - \mu(1-p_1)]\right]. \quad (23)$$

And in the case of $i = 10$:

$$g_1 = 10 \bullet \left(\frac{-1}{\mu}\right)\,\xi_9 \bullet f(\xi_9)$$

$$g_2 = 0$$

$$g_3 = -\left(\frac{RMI_{10}}{\mu}\right).$$

So

$$Asy.var\left(R\widehat{M}I_{10}\right) = g_1^2\,\sigma_{11} + g_3^2\,\sigma_{33} + 2g_1g_3\sigma_{13}$$

$$= \left[10\left(\frac{-1}{\mu}\right)\,\xi_9 \bullet f(\xi_9)\right]^2 \bullet \left[\frac{p_9(1-p_9)}{[f(\xi_9)]^2}\right] + \left[-\left(\frac{RMI_{10}}{\mu}\right)\right]^2 \bullet \sigma^2$$

$$+ 2\left[10\left(\frac{-1}{\mu}\right)\,\xi_9 \bullet f(\xi_9)\right]\left[-\left(\frac{RMI_{10}}{\mu}\right)\right] \bullet \left[\frac{[\xi_9 - \mu(1-p_9)]}{f(\xi_9)}\right]$$

$$= 10^2\left[\left(\frac{\xi_9}{\mu}\right)^2 p_9(1-p_9)\right] + \left(\frac{RMI_{10}}{\mu}\right)^2 \bullet \sigma^2$$

$$+ 2(10)\left[\left(\frac{\xi_9}{\mu}\right)\left(\frac{RMI_{10}}{\mu}\right)[\xi_9 - \mu(1-p_9)]\right]. \quad (24)$$

It then follows that

$$S.E.\left(R\widehat{M}I_i\right) = \left[\frac{Asy\widehat{.var}(R\widehat{M}I_i)}{N}\right]^{1/2}. \quad (25)$$

Note how both approaches lead to distribution-free asymptotic variances and standard errors, so that conventional statistical inference can be easily undertaken. Note also that the different approaches are not inconsistent, but only lead to different (and alternative) representations of the variance-covariance structure of the relative mean income estimates.

2.6 Application of QFA to Income Shares

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

$$IS_i \equiv \int_{R_i} \left(\frac{1}{\mu}\right) yf(y)dy \qquad \text{for } i = 1, \ldots, K, \tag{26}$$

with integration over the region R_i . But

$$IS_i = D_i \bullet \left(\frac{IS_i}{D_i}\right) = D_i \bullet RMI_i . \tag{27}$$

So IS_i is simply a given scalar proportion of RMI_i , and similarly,

$$I\hat{S}_i = D_i \bullet R\widehat{M}I_i . \tag{28}$$

Consequently,

$$Asy.var\left(I\hat{S}_i\right) = D_i^2 \bullet Asy.var\left(R\widehat{M}I_i\right) \tag{29}$$

and $$S.E.\left(I\hat{S}_i\right) = D_i \bullet S.E.\left(R\widehat{M}I_i\right) \tag{30}$$

for $i = 1, \ldots, K$.

If one goes back to first principles, one notes that

$$IS_i = N_i(\xi_{i-1}, \xi_i, \mu) \tag{31}$$

where N_i is the same integral function (19) in the last section. Thus applying the LWA results and Rao's linkage theorem to eq. (31) results in the same derivatives with respect to N_i and hence the same formulas – though rescaled by D_i – as in eqs. (22)-(24).

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

where now

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

So in the case of $i = 1$:

$$g_1 = 0$$

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

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

and

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

$$= \left(\frac{\xi_1}{\mu}\right)^2 p_1(1-p_1) + \left(\frac{IS_1}{\mu}\right)^2 \sigma^2 - 2\left(\frac{\xi_1}{\mu}\right)\left(\frac{IS_1}{\mu}\right)[\xi_1 - \mu(1-p_1)] . \qquad (32)$$

In the case of $i = 10$:

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

$$g_2 = 0$$

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

so

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

$$= \left(\frac{\xi_9}{\mu}\right)^2 p_9(1-p_9) + \left(\frac{IS_{10}}{\mu}\right)^2 \sigma^2 + 2\left(\frac{\xi_9}{\mu}\right)\left(\frac{IS_{10}}{\mu}\right)[\xi_9 - \mu(1-p_9)] . \qquad (33)$$

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

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

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

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

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

$$
\begin{aligned}
&= \left(\frac{\xi_{i-1}}{\mu}\right)^2 p_{i-1}(1-p_{i-1}) + \left(\frac{\xi_i}{\mu}\right)^2 p_i(1-p_i) + \left(\frac{IS_i}{\mu}\right)^2 \sigma^2 \\
&\quad - 2\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{\xi_i}{\mu}\right) p_{i-1}(1-p_i) \\
&\quad + 2\left(\frac{\xi_{i-1}}{\mu}\right)\left(\frac{IS_i}{\mu}\right)[\xi_{i-1} - \mu(1-p_{i-1})] \qquad (34) \\
&\quad - 2\left(\frac{\xi_i}{\mu}\right)\left(\frac{IS_i}{\mu}\right)[\xi_i - \mu(1-p_i)] .
\end{aligned}
$$

Note, incidentally, that just as IS_i is a ratio and hence units-free, so also is each term of its (asymptotic) variance and hence its standard error.

And, again, the standard error formulas for income shares are also distribution-free, and conventional statistical inference can be undertaken in straightforward fashion. Since we have not had to impose an assumption/restriction on the specific density functional form underlying the income distribution, this QFA approach can also be applied to highly skewed distributions as well, such as for wealth distributions.

3. Illustrative Empirical Results for Canada 1997-2015

3.1 Basic Data Sources and Sample Groups

The data used in this study come from the monthly Labour Force Survey microdata files (for May) from Statistics Canada for the period 1997 (when LFS microdata has become available) -2015. The variable of interest is individual worker's earnings. 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.

Summary statistics on the estimation samples – separately for male and female workers – appear in Appendix A, Table A1, at the end of the paper for selective years 1997, 2000, 2005, 2010, and 2015. The sample sizes are reasonably large – ranging from 23,175 (women workers in 1997) to 51,680 (men in 2015) – so that there should be considerable confidence in the (asymptotic-based) standard errors of the statistics reported. All dollar figures are expressed in 2002 constant dollars (based on the CPI deflator). As can be seen, overall mean (real) weekly earnings increased – with one exception (males between 2000 and 2005) – over pretty much the whole period. Between 1997 and 2015, males' average real weekly earnings ($\hat{\mu}$) went up by 10.0 percent and females' by 21.0 percent. But dispersion across earnings in the Canadian labour market also went up. The estimated standard deviation of earnings ($\hat{\sigma}$) rose by 18.0 percent for men and 25.4 percent for women. Note that weekly earnings is the product of hourly wage rates and hours worked in the survey week. So the higher growth figures for women reflect both higher wage rates and an increase in average hours worked by female workers in the Canadian labour market over the period covered.

It turns out from the asymptotic variance and standard error formulas of the previous section that σ / μ, the coefficient of variation of the earnings distribution, plays an important role in evaluating the confidence of many of the distributional statistics examined in this study. The third column of results in Table A1 shows that $\hat{\sigma} / \hat{\mu}$ figures for women are generally higher than for men in the Canadian labour market as women have a higher proportion of part-time workers, and they have also risen for both male and female workers over the 1997-2015 period.

When the observations in each sample are ordered by weekly earnings and decile earnings level cut-offs $(\hat{\xi}_i)$ calculated – again all in 2002 real earnings – the estimates are displayed in Appendix Table A2. These are the cut-off estimates on which all the quantile statistics and estimated asymptotic variance calculations are based. As can be seen, the 2001 recession did have a noticeable depressing effect on men's earnings up to and including the sixth decile level. But overall, the major story is the widespread (real) earnings increases experienced right across the earnings distributions over virtually the entire period. This is not the storyline often cited in the media – often based on United States results – and not the same as what happened in Canada over the 1980s and 1990s when many men experienced real earnings losses (Beach, 2016).

3.2 Earnings Shares Results

Shares of total earnings received by different quantile groups, separately for men and women, are presented in Table 1. The first column of the table provides the shares of the lowest-earning or bottom 10 percent of workers. The second columns includes the earnings shares of the bottom 20 percent. The third column lists that of the middle 60 percent of earners. The fourth column presents the share of the top 20 percent, and the last column does so for the top 10 percent. Note the overlap in coverage between the first two columns and between the last two columns so as to highlight the two ends of the distribution. Complete results on all five quintile earnings shares are also provided in Appendix C, Table C1, and further highlight the middle range of the distributions.

In both Tables 1 and C1, standard errors are included in parentheses for each earnings share statistic. Technical details on the standard error calculations are set out in Appendix B (which makes use of the general formulas in the previous section). Figures in square brackets are absolute values of (asymptotic) “t-ratios” of the estimated changes in earnings shares between 1997 and 2015.

As can be seen from Table 1, approximately 55 percent of total earnings are received by the middle 60 percent or broadly-speaking middle-class earners in the Canadian labour market. The bottom 20 percent of earners receive about 5.5-7.0 percent of total earnings, and the lowest 10 percent receive about 1.7-2.0 percent of labour market earnings. In contrast at the upper end of the distribution, the top 20 percent of earners receive about 36-40 percent of all earnings and the top-earning 10 percent of workers take home about 20-24 percent of all earnings. These results are pretty similar between men and women in the Canadian labour market.

In terms of changes over time, the broad result that is apparent is that the lower earnings shares show no statistically significant change over the 1997-2015 period as a whole, while the middle earnings share have declined statistically significantly and the upper earnings shares have significantly risen. A similar general pattern holds among both women and men in the labour market, but the pattern appears more marked among male earners than female earners. While this general pattern of change findings is not novel (see, for example, Beach, 2016), the results on the reliability or statistical significance of the finding is. This new result just serves to further reinforce our statistical understanding of what has been happening in the Canadian labour market over this period, and to highlight or focus attention on what observed distributional changes are indeed meaningful.

3.3 Quantile Mean Earnings Results

Table 2 displays quantile mean figures for the same five quantile groups as in the previous table. The full set of quintile means are provided in Appendix Table C2. Again, all figures are expressed in constant 2002 dollars. Here there are quite marked differences in quantile mean levels for men and women in the labour market. In the most recent year mean middle-class earnings for men was $769 versus $573 for women workers – again the figures reflect differences in both average hourly wage rates and average weekly hours worked on the job, and women are relatively more concentrated in lower-paying service-sector jobs while men are relatively more concentrated in manufacturing/construction/transportation sector jobs and higher-paying management and professional jobs. Mean earnings in 2015 for the lowest decile earnings group was $115 per week among female workers and $164 per week for male workers. At the upper end of the distribution, males in the top decile earned on average $1889 per week, while female workers took home $1496 a week on average.

The most salient feature of Table 2, however, is that the quantile mean figures all rose highly statistically significantly over the 1997-2015 period, so that real earnings gains were widely experienced right across the Canadian earnings distribution with larger gains experienced (in both dollar and percentage increase terms) by female earners in the labour market. This finding of widely shared real-earnings gains over the period as a whole is consistent with the results already noted for the quantile cut-off levels in Appendix Table A2. The statistical significance of the gains was most marked for the broad middle-class group of workers over the 1997-2015 period under review.

3.4 Relative Mean Earnings Results

Relative mean earnings rates by quantile group are presented in Table 3. Each entry is the ratio of the corresponding quantile mean to the overall mean earnings level. Similar quintile ratios appear in Appendix Table C3. Relative mean rates for the middle 60 percent of workers range from 0.90 to 0.96. For the bottom decile of workers, they vary from 0.17 to 0.20, and for the top decile group from 2.15-2.27. Female workers have slightly higher rates than males among the top 20 percent of workers and slightly lower rates over the lower 80 percent.

Over the 1997-2015 period, slight but highly statistically significant relative losses occurred among the middle earnings group (by about 0.02 ratio points), while statistically significant relative gains occurred over the upper regions of the distributions (of about 0.06-0.09 ratio points). From the results in Table C3, one can see that the relative losses are most extreme among the third quintile (i.e., middlemost 20 percent) of male earners. The relative gains among higher earners are essentially concentrated in the top 10 percent earnings group (by about 0.09 ratio points) for both male and female workers.

The quantile mean results already seen in Table 2 can be thought of as a welfare indicator and the product of two key dimensions. One is the economic efficiency dimension or overall mean earnings level across all workers in an economy. This is represented by $\hat{\mu}$ figures noted earlier in Appendix Table A1. The second is an equity dimension as represented by the relative mean earnings rates in Table 3. Overall means went up by 10.0 percent among male workers and by 21.0 percent among female workers over the 1997-2015 period. When measured over almost a twenty-year interval, efficiency gains (via productivity advances and demographic evolution) contribute a great deal to overall economic well-being. However, evidently the gains were not evenly shared. As is evident, middle earners lost out relatively whereas the top earners were the big winners over the period. A rising tide indeed raised all boats to recall a phrase, but evidently some boats moved up much higher than others.

3.5 Results on Earnings Gaps and Earnings Differentials

Table 4 results are displayed in a somewhat different format. This table focuses on earnings gaps and earnings differentials. Results in column 1 refer to the gap in quantile mean earnings levels between that of the middle quintile ($\hat{\mu}_{q3}$) and that of the lowest decile ($\hat{\mu}_1$) . Column 2 results show the gap between the top decile mean ($\hat{\mu}_{10}$) and the middle quintile mean ($\hat{\mu}_{q3}$) . Again these are both expressed in 2002 dollars. The last two columns highlight these earnings differences in percentage or relative terms. Column 3 indicates the lower earnings ratio of $\hat{\mu}_{q3}$ to $\hat{\mu}_1$ and the last column shows the upper earnings ratio $\hat{\mu}_{10}$ / $\hat{\mu}_{q3}$. In dollar terms all the earnings gaps widened over the 1997-2015 period as a whole with three of the four increases highly statistically significant. In relative or percentage terms, though, the lower earnings differential narrowed (but not statistically significantly), while the upper earnings differential widened highly statistically significantly. Again, this reinforces the findings in the previous tables of the middle earners losing out relatively but statistically significantly to the top earners in the Canadian labour market over this period.

4. Overview, Findings and Conclusions

Recent major policy initiatives to deal with the COVID-19 pandemic and on-going major labour market developments reflecting changing patterns of globalization, automation and demographics argue for the need for statistical techniques that allow detailed analysis of disaggregative distributional change. Such a tool box set of disaggregative inequality measures was forwarded in Beach (2021) which also developed a statistical methodology (a quantile function approach or QFA) that enables the calculation of relatively simple standard error formulas for these measures, thus allowing one to perform formal statistical inference with these measures. The present paper applies and extends this approach to percentile-based inequality statistics (such as decile income shares and quintile mean income levels) typically published by official statistical agencies such as Statistics Canada and the U.S. Bureau of the Census. In so doing, it shows that the resulting standard error formulas turn out to be distribution-free and are thus relatively simple and easy to implement. The paper then illustrates the use of the tool box set of distributional measures and the corresponding standard error formulas with an examination of workers' earnings in the Canadian labour market with Labour Force Survey micro data over the period 1997-2015.

The paper highlights four sets of disaggregative tool box measures characterizing the distribution of earnings and how it has changed over this period:

- earnings shares
- quantile means
- quantile earnings gaps and earnings differentials, and
- relative mean earnings rates.

Specific standard error formulas for these tool box measures are developed in Section 2.

Two general findings arise from the empirical results in Tables 1-4 of the paper. First, gains in workers' earnings were very widespread over the 1997-2015 period, so that all quantile groups experienced statistically significant increases in real earnings levels. Not only did the overall earnings pie increase, but the size of the pie slice going to each quantile group went up too.

However, second, the distribution of the earnings gains was uneven, with the middle-class earnings shares falling and the top earnings shares going up highly statistically significantly, so that the upper earnings differential widened very significantly and top earners pulled away from the rest of the earnings distribution. While each pie slice of earnings was getting bigger over the 1997-2015 period, the gains in slice size disproportionately occurred among the top earners in the labour market. To further mix metaphors, what occurred was not a rising tide as much as a ski-lift experience at a ski resort where the skilled-workers' ski lift run went much higher up the mountain than the beginners' ski lift run which ended only part-way up the slope before the run became quite steep.

This paper has shown that calculating standard errors for a whole tool box of disaggregative inequality measures is straightforward and quite easy to do. It would thus be useful if the statistical agencies that provide official income share and quantile mean income data would also include either the accompanying standard error reliability estimates that go with them – if only for the most recent year's estimates perhaps and provided in the accompanying documentation on the source survey's methodology. This is the "nifty fix" referred to in the title of the paper. Failing that, it would be very helpful if the data provider included accompanying information so that users and practitioners can calculate the standard error reliability indicators themselves. This would mean providing information on the actual estimation sample sizes (in addition to overall survey samples and response rates) and on the standard deviation ($\hat{\sigma}$) or coefficient of variation ($\hat{\sigma}$ / $\hat{\mu}$) for the estimation samples on which the income shares and quantile mean estimates are based.

Table 1

Selected Quantile Earnings Shares of All Workers Age 25-59, Canada, 1997-2015

LFS Data on Usual Weekly Earnings

(percent)

	Bottom 10%	Bottom 20%	Middle 60%	Top 20%	Top 10%
			Males		
1997	1.9942 (.06857)	6.6092 (.1364)	56.7702 (.3948)	36.6206 (.3819)	20.5183 (.3311)
2000	2.0215 (.06891)	7.5696 (.1404)	55.9830 (.3802)	36.4474 (.3674)	21.5057 (.3228)
2005	2.0101 (.06640)	6.4217 (.1310)	56.0979 (.3883)	37.4804 (.3775)	21.8529 (.3312)
2010	1.9611 (.06352)	6.3754 (.1272)	55.6572 (.3846)	37.9675 (.3749)	22.2446 (.3318)
2015	1.9675 (.04531)	6.4466 (.09150)	55.2459 (.2794)	38.3075 (.2726)	22.6621 (.2428)
Change 1997-2015	-.0267 [0.32]	-.1626 [0.99]	-1.5243 [-3.15]	1.6869 [3.60]	2.1438 [5.22]
			Females		
1997	1.7450 (.05673)	5.4735 (.1243)	55.5242 (.4248)	39.0023 (.4160)	22.5901 (.3722)
2000	1.9016 (.05767)	5.8430 (.1267)	56.0789 (.4202)	38.0780 (.4098)	22.5539 (.3630)
2005	1.8642 (.05570)	5.7073 (.1199)	54.6469 (.4043)	39.6458 (.3965)	23.4495 (.3581)
2010	1.7792 (.05147)	5.5704 (.1151)	54.7200 (.3918)	39.7096 (.3846)	23.4134 (.3517)
2015	1.8214 (.0378)	5.6208 (.08542)	54.2498 (.2845)	40.1294 (.2797)	23.4761 (.2582)
Change 1997-2015	.0764 [1.12]	.1473 [0.98]	-1.2744 [2.49]	1.1271 [2.25]	.8860 [1.96]

Source: Based on Statistics Canada's PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 2

Selected Quantile Mean Earnings of All Workers Age 25-59, Canada, 1997-2015

LFS Data on Usual Weekly Earnings

(real 2002 $)

	Bottom 10%	Bottom 20%	Middle 60%	Top 20%	Top 10%
			Males		
1997	150.91 (5.162)	247.87 (5.082)	717.24 (4.417)	1391.81 (13.57)	1649.47 (24.44)
2000	154.74 (5.249)	266.60 (5.278)	734.21 (4.309)	1404.17 (13.19)	1646.79 (24.04)
2005	152.35 (5.038)	245.06 (4.920)	713.54 (4.374)	1431.25 (13.46)	1686.51 (24.54)
2010	159.74 (5.146)	259.38 (5.101)	755.78 (4.621)	1552.70 (14.23)	1813.63 (26.24)
2015	163.85 (3.749)	264.58 (3.744)	769.24 (3.421)	1604.44 (10.54)	1888.89 (19.60)
Change 1997-2015	12.94 [2.03]	16.71 [2.65]	52.00 [9.31]	212.63 [12.37]	239.42 [7.64]
			Females		
1997	90.558 (2.946)	143.02 (3.201)	484.57 (3.259)	1020.66 (10.07)	1190.51 (18.88)
2000	99.427 (3.066)	155.37 (3.339)	498.50 (3.309)	1046.07 (10.22)	1213.88 (18.87)
2005	102.83 (3.086)	159.00 (3.292)	509.73 (3.293)	1108.68 (10.18)	1313.38 (19.33)
2010	107.18 (3.115)	169.88 (3.453)	556.30 (3.489)	1217.62 (10.78)	1429.82 (20.77)
2015	115.12 (2.372)	177.79 (2.659)	572.84 (2.625)	1271.26 (8.101)	1496.16 (15.81)
Change 1997-2015	24.56 [6.49]	34.77 [8.36]	88.27 [21.09]	250.60 [19.38]	305.65 [12.41]

Source: Based on Statistics Canada's PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 3

Relative Mean Earnings for Selected Quantiles of All Workers Age 25-59, Canada, 1997-2015

LFS Data on Usual Weekly Earnings

	Bottom 10%	Bottom 20%	Middle 60%	Top 20%	Top 10%
			Males		
1997	0.19939 (.006857)	0.32750 (.006819)	0.94766 (.006579)	1.83895 (.01909)	2.17939 (.03311)
2000	0.20209 (.006891)	0.34818 (.007019)	0.95889 (.006336)	1.83386 (.01837)	2.15073 (.03228)
2005	0.19960 (.006640)	0.32106 (.006548)	0.93482 (.006472)	1.87511 (.01887)	2.20953 (.03312)
2010	0.19596 (.006352)	0.31819 (.006362)	0.92714 (.006410)	1.90476 (.01874)	2.22485 (.03317)
2015	0.19672 (.004531)	0.31766 (.004575)	0.92356 (.004656)	1.92631 (.01363)	2.26782 (.02428)
Change 1997-2015	-.00267 [0.32]	-.00984 [1.20]	-0.2410 [2.99]	.08736 [3.72]	.08843 [2.15]
			Females		
1997	0.17303 (.005673)	0.27327 (.006215)	0.92587 (.007080)	1.95017 (.02080)	2.27470 (.03722)
2000	0.18544 (.005767)	0.28977 (.006333)	0.92973 (.007003)	1.95097 (.02049)	2.26394 (.03630)
2005	0.18392 (.005570)	0.28439 (.005996)	0.91171 (.006739)	1.98301 (.01983)	2.34914 (.03581)
2010	0.17552 (.005147)	0.27820 (.005754)	0.91103 (.006530)	1.99404 (.01923)	2.34155 (.03517)
2015	0.18176 (.003779)	0.28070 (.004271)	0.90442 (.004741)	2.00710 (.01399)	2.36218 (.02582)
Change 1997-2015	.00873 [1.28]	.00743 [0.99]	-.02145 [2.52]	.05693 [2.27]	.08748 [1.93]

Source: Based on Statistics Canada's PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table 4

Quantile Mean Earnings Gaps and Differentials for All Workers Age 25-59, Canada, 1997-2015

LFS Data on Usual Weekly Earnings

(dollar values in real 2002 $)

	Q3-D1 Gap ($)	D10-Q3 Gap ($)	Q3/D1 Differential (ratio)	D10/Q3 Differential (ratio)
		Males		
1997	564.11 (11.06)	934.45 (28.47)	4.7381 (.1770)	2.3069 (.05161)
2000	566.92 (11.00)	925.13 (27.99)	4.6637 (.1726)	2.2820 (.05023)
2005	550.88 (10.72)	983.28 (28.50)	4.6159 (.1668)	2.3982 (.05303)
2010	580.86 (11.08)	1073.03 (30.30)	4.6363 (.1625)	2.4489 (.05360)
2015	586.19 (8.068)	1138.85 (22.62)	4.5776 (.1144)	2.5184 (.03987)
Change 1997-2015	22.08 [1.61]	204.40 [5.62]	-.16046 [0.76]	.21150 [3.24]
		Females		
1997	389.27 (7.359)	710.68 (21.73)	5.2986 (.1894)	2.4811 (.05919)
2000	387.40 (7.466)	727.05 (21.78)	4.8964 (.1670)	2.4934 (.05880)
2005	400.59 (7.444)	809.96 (22.16)	4.8957 (.1625)	2.6089 (.05848)
2010	439.03 (7.749)	883.61 (23.76)	5.0962 (.1634)	2.6177 (.05758)
2015	451.13 (5.830)	929.91 (18.01)	4.9188 (.1125)	2.6422 (.04202)
Change 1997-2015	61.86 [6.59]	219.23 [7.77]	-.37981 [1.72]	.16112 [2.22]

Source: See Tables 2 and C2.
Figures in parentheses are (asymptotic) standard errors.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

References

Beach, C.M. (2021) “A Useful Empirical Tool Box for Distributional Analysis”, Queen’s University, Department of Economics Discussion Paper No. 1466, September.

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.

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

Appendix A

Table A1

Summary Statistics on Canadian Weekly Earnings for All Workers Age 25-59

Selective Years, 1997-2015

(real 2002 dollars)

	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$ / $\hat{\mu}$	**NOBS**
		Males		
1997	425.20	756.85	0.56180	24,615
2000	427.01	765.69	0.55768	25,511
2005	441.03	763.29	0.57780	25,831
2010	478.63	815.17	0.58715	26,621
2015	501.77	832.91	0.60243	51,680
		Females		
1997	322.54	523.37	0.61628	23,175
2000	327.11	536.18	0.61007	23,917
2005	353.74	559.09	0.63271	25,414
2010	387.40	610.63	0.63443	27,422
2015	404.54	633.38	0.63870	51,658

Note: Based on May Labour Force Surveys.

Table A2

Decile Cut-Offs on Canadian Weekly Earnings for All Workers Age 25-59

Selective Years, 1997-2015

(real 2002 dollars)

	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{\xi}_3$	$\hat{\xi}_4$	$\hat{\xi}_5$	$\hat{\xi}_6$	$\hat{\xi}_7$	$\hat{\xi}_8$	$\hat{\xi}_9$
Males									
1997	269.93	398.67	511.63	620.16	708.75	809.24	916.94	1064.73	1277.96
2000	279.45	421.50	523.71	628.03	716.54	810.62	927.29	1053.74	1279.77
2005	269.92	395.38	513.59	605.44	702.91	808.34	931.58	1081.49	1315.84
2010	279.88	416.17	533.10	637.15	742.91	851.57	992.22	1160.79	1427.34
2015	284.08	425.53	535.86	630.42	756.50	873.13	1008.67	1197.79	1485.11
Females									
1997	149.50	243.63	326.25	398.67	476.63	553.71	642.30	766.67	958.14
2000	158.06	258.17	334.46	405.48	786.04	569.02	663.86	790.31	972.81
2005	164.01	262.42	337.39	421.74	504.59	581.07	679.33	811.05	1027.30
2010	171.97	285.90	369.73	451.42	546.86	635.80	744.20	892.86	1146.41
2015	179.67	302.13	380.17	472.81	562.65	651.54	758.08	921.36	1197.73

Note: Based on May Labour Force Surveys.

Appendix B

Statistical Inference Formulas Used to Calculate Standard Errors for Tables 1-4

Because Tables 1-4 involve mixed income group sizes over different regions of a distribution, it may be useful to indicate exactly the formulas used to calculate the different (asymptotic) variances and corresponding standard errors. Since

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

for any statistic, stat, the standard errors depend on the size of the estimation samples, these are provided for each sample in Appendix A. Thus in this appendix, we focus just on the relevant (asymptotic) variance formulas.

For Table 1 on Earnings Shares

In the case of the bottom 10 percent share, use the formula in eq. (31) in the text with $p_1 = 0.10$ and ξ_1 is the first decile earnings cut-off (see Table A2 values), and IS_1 is the bottom decile earnings share.

In the case of the bottom 20 percent or quintile share, again use eq. (31), but now with $p_1 = 0.20$ and ξ_1 is the second decile earnings cut-off (i.e., the first quintile earnings cut-off) and IS_1 is the bottom quintile earnings share. That is,

$$Asy.var\left(I\hat{S}_1\right) = p_1(1-p_1)\left(\frac{\xi_1}{\mu}\right)^2 + (IS_1)^2 \bullet \left(\frac{\sigma}{\mu}\right)^2$$

$$Asy.var\left(I\hat{S}_{q1}\right) = p_2(1-p_2)\left(\frac{\xi_2}{\mu}\right)^2 + \left(IS_{q1}\right)^2 \bullet \left(\frac{\sigma}{\mu}\right)^2.$$

In the case of the middle 60 percent earnings share, use the formula in eq. (33) with $p_{i-1} = 0.20$, $p_i = 0.80$, ξ_{i-1} is the second decile earnings cut-off (i.e., the cut-off value at the lower end of the 60 percent quantile interval), ξ_i is the eighth decile earnings cut-off (i.e., the cut-off value at the upper end of the mid 60 percent interval), and IS_i is the sum of the middle three quintile earnings shares. That is,

$$Asy.var(I\hat{S}_M) = p_2(1-p_2)\left(\frac{\xi_2}{\mu}\right)^2 + p_8(1-p_8)\left(\frac{\xi_8}{\mu}\right)^2 + (IS_M)^2 \bullet \left(\frac{\sigma}{\mu}\right)^2$$

$$-2\left(\frac{\xi_2}{\mu}\right)\left(\frac{\xi_8}{\mu}\right)p_2(1-p_8)$$

$$+2\left(\frac{\xi_2}{\mu}\right) \bullet IS_M\left[\left(\frac{\xi_2}{\mu}\right) - (1-p_2)\right]$$

$$-2\left(\frac{\xi_8}{\mu}\right) \bullet IS_M\left[\left(\frac{\xi_8}{\mu}\right) - (1-p_8)\right].$$

In the case of the top 20 percent or quintile share, use eq. (32) with , $p_9 = 0.80$, ξ_9 refers to the next-to-top decile earnings cut-off (i.e., the top quintile earnings cut-off), and IS_{10} refers to the top quintile earnings share.

In the case of the top 10 percent share, again use eq. (32) with $p_9 = 0.90$, ξ_9 now referring to the top decile earnings cut-off, and IS_{10} indicating the top decile earnings share. That is,

$$Asy.var(I\hat{S}_{10}) = p_9(1-p_9)\left(\frac{\xi_9}{\mu}\right)^2 + \left(I\hat{S}_{10}\right)^2 \bullet \left(\frac{\sigma}{\mu}\right)^2$$

$$Asy.var(I\hat{S}_{q5}) = p_8(1-p_8)\left(\frac{\xi_8}{\mu}\right)^2 + \left(I\hat{S}_{q5}\right)^2 \bullet \left(\frac{\sigma}{\mu}\right)^2.$$

<u>For Table 3 on Relative Mean Earnings Ratios</u>

Since $Asy.var(I\hat{S}_i) = D_i^2 \bullet Asy.var(R\widehat{M}I_i)$ for each quantile group *i* – see eq. (29) – it follows that

$$Asy.var(R\widehat{M}I_i) = \left(\frac{1}{D_i^2}\right) \bullet Asy.var(I\hat{S}_i)\,. \tag{c1}$$

So, instead of calculating all the (asymptotic) variance terms for Table 3 from first principles, simply use the (asymptotic) variance estimates for Table 1, and make the proportional adjustment indicated in eq. (c1) (i.e., divide S.E. figures in Table 1 by D_i) where

D_i = 0.10 for col. 1

D_i = 0.20 for col. 2

D_i = 0.60 for col. 3

D_i = 0.20 for col. 4

and D_i = 0.10 for col. 5 .

For Table 2 on Quantile Mean Earnings

In the case of the bottom 10 percent mean (col. 1) and the bottom 20 percent mean (col. 2), use the formula

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

For the bottom 10 percent $D_1 = 0.10$, $p_1 = 0.10$, and ξ_1 is the first decile cut-off earnings level. For the bottom 20 percent, μ_1 refers to the first quintile mean, $D_1 = 0.20$, $p_1 = 0.20$, ξ_1 is the second decile (i.e., first quintile) cut-off earnings level. That is,

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

$$Asy.var(\hat{\mu}_{q1}) = \left(\frac{1}{p_2}\right)^2 p_2(1 - p_2)\,\xi_2^2 .$$

In the case of the middle 60 percent mean (col. 3), use the formula

$$Asy.var(\hat{\mu}_i) = \left(\frac{1}{D_i}\right)^2 [p_2(1 - p_2)\,\xi_2^2 + p_8(1 - p_8)\,\xi_8^2 - 2\,p_2(1 - p_8)\,\xi_2\,\xi_8] \quad \text{(c2)}$$

where now $D_i = p_8 - p_2 = 0.60$, the range of this quantile interval runs from a lower cut-off earnings level of ξ_2 (the second decile cut-off) to an upper cut-off earnings level of ξ_8 (the eighth decile cut-off), and the proportion covered by this range goes from $p_2 = 0.20$ up to $p_8 = 0.80$ (hence the D_i value of 0.80 – 0.20 = 0.60).

In the case of the top 20 percent mean (col. 4) and top 10 percent mean (col. 5), use the formula

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

For the top 20 percent, μ_K refers to mean of the top quintile (where K = 5), ξ_{K-1} is the fourth quintile cut-off level (i.e., the eighth decile cut-off level), and $p_{K-1} = p_8 = 0.80$, so $D_K = 0.20$. For the top 10 percent, μ_K refers to the mean of the top or tenth decile (where K = 10), ξ_{K-1} is the ninth decile cut-off level (i.e., ξ_9), and $p_{K-1} = p_9 = 0.90$, so $D_K = 0.10$. that is,

$$Asy.var(\hat{\mu}_{10}) = \left(\frac{1}{1-p_9}\right)^2 p_9(1 - p_9)\,\xi_9^2$$

$$Asy.var(\hat{\mu}_{q5}) = \left(\frac{1}{1-p_8}\right)^2 p_8(1 - p_8)\,\xi_8^2 .$$

These formulas are all variations of eqs. (8a)-(8c) in section 2.3 of the text.

For Table 4 on Mean Earnings Gaps and Differentials

Here the complication is that decile means are being compared to the mean of the middle quintile. Let the decile means be designated by μ_1 and μ_{10} and middle quintile mean by μ_Q. Again for the first decile mean, use

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

where $D_1 = 0.10$ and $p_1 = 0.10$. For the top decile mean, use

$$Asy.var(\hat{\mu}_{10}) = \left(\frac{1}{D_{10}}\right)^2 p_9(1 - p_9)\,\xi_9^2$$

where $D_{10} = 0.10$ and $p_9 = 0.90$.

Now $\hat{\mu}_Q = \left(\frac{1}{2}\right)(\hat{\mu}_5 + \hat{\mu}_6)$

and $\mu_Q = \left(\frac{1}{2}\right)(\mu_5 + \mu_6)$. So

$$Cov\left(\hat{\mu}_1, \hat{\mu}_Q\right) = E\left[(\hat{\mu}_1 - \mu_1)\left(\hat{\mu}_Q - \mu_Q\right)\right]$$

which by substitution for $\hat{\mu}_Q$ and μ_Q leads to

$$Cov\left(\hat{\mu}_1, \hat{\mu}_Q\right) = \left(\frac{1}{2}\right) Cov(\hat{\mu}_1, \hat{\mu}_5) + \left(\frac{1}{2}\right) Cov(\hat{\mu}_1, \hat{\mu}_6)\,. \quad \text{(c3)}$$

Similarly,

$$Cov\left(\hat{\mu}_{10}, \hat{\mu}_Q\right) = \left(\frac{1}{2}\right) Cov(\hat{\mu}_{10}, \hat{\mu}_5) + \left(\frac{1}{2}\right) Cov(\hat{\mu}_{10}, \hat{\mu}_6)\,. \quad \text{(c4)}$$

Since the covariance result holds for all *N*, it also holds asymptotically for *Asy.cov* expressions corresponding to (c3) and (c4). Also

$$Asy.var\left(\hat{\mu}_Q\right) = \left(\frac{1}{0.20}\right)^2 \left[p_4(1 - p_4)\,\xi_4^2 + p_6(1 - p_6)\,\xi_6^2 \right.$$
$$\left. - 2p_4(1 - p_6)\xi_4\,\xi_6\right] \quad \text{(c5)}$$

analogous to applying (c2) to quintiles (where $K_i = 0.20$) with lower cut-off of ξ_4 and upper cut-off of ξ_6 .

Now from eqs. (10b) and (10c) of section 2.4 of the text, one can see that:

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_5) = \left(\frac{1}{0.10}\right)^2 \left[-p_1(1 - p_4)\xi_1\,\xi_4 + p_1(1 - p_5)\xi_1\,\xi_5\right]$$

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_6) = \left(\frac{1}{0.10}\right)^2 \left[-p_1(1 - p_5)\xi_1\,\xi_5 + p_1(1 - p_6)\xi_1\,\xi_6\right] \quad \text{(c6)}$$

$$Asy.cov(\hat{\mu}_5, \hat{\mu}_{10}) = \left(\frac{1}{0.10}\right)^2 \left[p_4(1 - p_9)\xi_4\,\xi_9 - p_5(1 - p_9)\xi_5\,\xi_9\right]$$

$$Asy.cov(\hat{\mu}_6, \hat{\mu}_{10}) = \left(\frac{1}{0.10}\right)^2 \left[p_5(1 - p_9)\xi_5\,\xi_9 - p_6(1 - p_9)\xi_6\,\xi_9\right].$$

For the mean earnings gaps in cols. 1 and 2 of Table 4, then, use eq. (12):

$$Asy.var\left(\hat{\mu}_{10} - \hat{\mu}_Q\right) = Asy.var(\hat{\mu}_{10}) + Asy.var\left(\hat{\mu}_Q\right) - 2\,Asy.cov\left(\hat{\mu}_{10}, \hat{\mu}_Q\right) \quad \text{(c7)}$$

$$Asy.var(\hat{\mu}_Q - \hat{\mu}_1) = Asy.var(\hat{\mu}_Q) + Asy.var(\hat{\mu}_1) - 2\,Asy.cov(\hat{\mu}_1, \hat{\mu}_Q)\ .$$

And for the mean earnings differentials in cols. 3 and 4 of Table 4, use eq. (14):

$$Asy.var(\hat{\mu}_{10} \,/\, \hat{\mu}_Q)$$

$$= \left(\frac{-\mu_{10}}{\mu_Q^2}\right)^2 \bullet Asy.var(\hat{\mu}_Q) + \left(\frac{1}{\mu_Q}\right)^2 \bullet Asy.var(\hat{\mu}_{10})$$

$$-2\left(\frac{\mu_{10}}{\mu_Q^2}\right)\left(\frac{1}{\mu_Q}\right) \bullet Asy.cov(\hat{\mu}_Q, \hat{\mu}_{10}) \qquad \text{(c8)}$$

$$Asy.var(\hat{\mu}_Q \,/\, \hat{\mu}_1)$$

$$= \left(\frac{-\mu_Q}{\mu_1^2}\right)^2 \bullet Asy.var(\hat{\mu}_1) + \left(\frac{1}{\mu_1}\right)^2 \bullet Asy.var(\hat{\mu}_Q)$$

$$-2\left(\frac{\mu_Q}{\mu_1^2}\right)\left(\frac{1}{\mu_1}\right) \bullet Asy.cov(\hat{\mu}_1, \hat{\mu}_Q)\ . \qquad \text{(c8)}$$

The sequence of calculations then is straightforward. First, calculate all the required (asymptotic) variances for $\hat{\mu}_1$, $\hat{\mu}_{10}$, and $\hat{\mu}_Q$. Then compute the four (asymptotic) covariances in eq. (c6), and use these to calculate the (asymptotic) covariances

$$Asy.cov(\hat{\mu}_1, \hat{\mu}_Q) = (0.5)\,Asy.cov(\hat{\mu}_1, \hat{\mu}_5) + (0.5)\,Asy.cov(\hat{\mu}_1, \hat{\mu}_6)$$

$$Asy.cov(\hat{\mu}_Q, \hat{\mu}_{10}) = (0.5)\,Asy.cov(\hat{\mu}_5, \hat{\mu}_{10}) + (0.5)\,Asy.cov(\hat{\mu}_6, \hat{\mu}_{10})\ .$$

Then plug in these expressions on the right-hand side of eqs. (c7) for the mean earnings gaps and of eqs. (c8)-(c9) for the mean earnings differentials, and replace all unknowns by their sample estimates in order to compute the standard errors reported in Table 4.

<u>Practical Concerns</u>

In the calculation/programming of the formulas in this paper, several practical concerns should usefully be kept in mind.

In the Table 1 calculations of $Asy.var(\hat{IS_i})$, note that the p_i and IS_i terms need to be on the same scale. If the p_i's are expressed as proportions (e.g., $p_1 = 0.10$), then so also should the IS_i's. This means dividing the reported IS_i figures in Table 1 by 100. Second, note that the relative importance of the separate terms can vary dramatically. In the asymptotic variance formulas for IS_1, and IS_{q1}, the ξ_i component accounts typically for 95 percent or more of the total variation; and for IS_{10} and IS_{q5}, about 90 percent or so. In the asymptotic variance formula for IS_M, however, the dominant component is ξ_8 or the second term followed by the σ^2 or third term. The last two terms, while elegant, account for only a very small amount of the total variation in the formula.

In the Table 2 asymptotic variance calculations for $\hat{\mu}_M$, again the ξ_8 component or second term is very much the dominant term, about six times the size of the next closest (or covariance) term.

In the Table 3 calculations, again note the need for proper scaling. Standard errors are always in the same units of the statistic they are attached to – so their implied t-ratios are units-free. In Table 3, the RMI_i ratios are expressed in proportions, while in Table 1 the income shares are reported in percentages. So one needs to rescale the reported standard errors of IS_i statistics by dividing the Table 1 standard errors by 100 and then divide these rescaled standard errors by the appropriate D_i values.

Appendix C

Table C1

Quintile Earnings Shares of All Workers Age 25-29, Canada, 1997-2015

LFS Data on Usual Weekly Earnings

(percent)

	Q1	Q2	Q3	Q4	Q5
Males					
1997	6.6092 (.1364)	14.1542 (.2088)	18.0713 (.2577)	24.5447 (.3150)	36.6206 (.3819)
2000	7.5696 (.1404)	12.6649 (.2036)	18.8193 (.2517)	24.4989 (.3045)	36.4474 (.3674)
2005	6.4217 (.1310)	13.3178 (.1972)	18.4371 (.2494)	24.3430 (.3072)	37.4804 (.3775)
2010	6.3754 (.1272)	13.0844 (.1916)	18.1447 (.2426)	24.4281 (.3032)	37.9675 (.3749)
2015	6.4466 (.09150)	12.9440 (.1333)	17.6812 (.1741)	24.6206 (.2197)	38.3075 (.2726)
Change 1997-2015	-.1626 [0.99]	-1.2102 [4.89]	-.3901 [1.25]	.0759 [0.20]	1.6869 [3.60]
Females					
1997	5.4735 (.1243)	12.8550 (.2018)	17.7659 (.2627)	24.9033 (.3335)	39.0023 (.4160)
2000	5.8430 (.1266)	12.3921 (.1961)	18.2362 (.2594)	25.4507 (.3301)	38.0780 (.4098)
2005	5.7073 (.1199)	12.4753 (.1907)	17.7166 (.2474)	24.4550 (.3152)	39.6458 (.3965)
2010	5.5704 (.1151)	12.2108 (.1798)	17.8374 (.2384)	24.6719 (.3023)	39.7096 (.3846)
2015	5.6208 (.08542)	12.9663 (.1326)	16.9934 (.1715)	24.2900 (.2203)	40.1294 (.2798)
Change 1997-2015	.1473 [0.98]	.1113 [0.46]	-.7725 [2.46]	-.6133 [1.53]	1.1271 [2.25]

Source: Based on Statistics Canada’s PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors.
Figures in square brackets are absolute (asymptotic) “t-ratios”.

Table C2

Quintile Mean Earnings of All Workers Age 25-29, Canada, 1997-2015

LFS Data on Usual Weekly Earnings

(real 2002 $)

	Q1	Q2	Q3	Q4	Q5
Males					
1997	247.87 (5.082)	517.48 (7.701)	715.02 (9.501)	925.33 (11.568)	1391.81 (13.573)
2000	266.60 (5.278)	529.89 (7.640)	721.66 (9.363)	932.53 (11.303)	1404.17 (13.195)
2005	245.06 (4.920)	507.34 (7.331)	703.23 (9.238)	930.84 (11.394)	1431.25 (13.458)
2010	259.38 (5.101)	533.41 (7.600)	740.60 (9.586)	992.52 (11.963)	1552.70 (14.229)
2015	264.58 (3.744)	536.82 (5.386)	750.04 (7.032)	1020.20 (8.840)	1604.44 (10.538)
Change 1997-2015	16.71 [2.65]	19.34 [2.06]	35.02 [2.96]	94.87 [6.52]	212.63 [12.37]
Females					
1997	143.02 (3.201)	325.92 (5.123)	479.83 (6.658)	653.33 (8.422)	1020.66 (10.073)
2000	155.37 (3.339)	333.20 (5.112)	486.83 (6.730)	672.49 (8.536)	1046.07 (10.221)
2005	159.00 (3.292)	343.18 (5.167)	503.42 (6.676)	685.74 (8.484)	1108.68 (10.175)
2010	169.88 (3.453)	371.92 (5.317)	546.21 (7.022)	750.45 (8.973)	1217.62 (10.783)
2015	177.79 (2.659)	390.01 (4.054)	566.25 (5.250)	772.44 (6.730)	1271.26 (8.108)
Change 1997-2015	34.77 [8.36]	64.09 [9.81]	86.42 [10.19]	119.11 [11.05]	250.60 [19.38]

Source: Based on Statistics Canada's PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors.
Figures in square brackets are absolute (asymptotic) "t-ratios".

Table C3

Quintile Relative Mean Earnings for All Workers Age 25-29, Canada, 1997-2015

LFS Data on Usual Weekly Earnings

	Q1	Q2	Q3	Q4	Q5
			Males		
1997	0.32750 (.006819)	0.68373 (.01044)	0.94473 (.01289)	1.22261 (.01575)	1.83895 (.01909)
2000	0.34818 (.007019)	0.69204 (.01018)	0.94250 (.01259)	1.21789 (.01522)	1.83386 (.01837)
2005	0.32106 (.006548)	0.66468 (.009859)	0.92131 (.01247)	1.21951 (.01536)	1.87511 (.01887)
2010	0.31819 (.006362)	0.65435 (.009578)	0.90852 (.01213)	1.21756 (.01516)	1.90476 (.01874)
2015	0.31766 (.004575)	0.64451 (.006665)	0.90051 (.008706)	1.22486 (.01098)	1.92631 (.01363)
Change 1997-2015	-.00984 [1.20]	-.03922 [3.17]	-.04422 [2.84]	.00225 [0.12]	.08736 [3.72]
			Females		
1997	0.27327 (.006215)	0.62273 (.01009)	0.91681 (.01314)	1.24831 (.01668)	1.95017 (.02080)
2000	0.28977 (.006333)	0.62143 (.009806)	0.90796 (.01297)	1.25422 (.01651)	1.95097 (.02049)
2005	0.28439 (.005996)	0.61382 (.009534)	0.90043 (.01237)	1.22653 (.01576)	1.98301 (.01983)
2010	0.27820 (.005754)	0.60908 (.008987)	0.89450 (.01192)	1.22898 (.01511)	1.99404 (.01923)
2015	0.28070 (.004271)	0.61576 (.006631)	0.89401 (.008575)	1.21955 (.01101)	2.00710 (.01399)
Change 1997-2015	.00743 [0.99]	-.00697 [0.58]	-.02280 [1.45]	-.02876 [1.44]	.05693 [2.27]

Source: Based on Statistics Canada's PUMF files for May Labour Force Surveys.
Figures in parentheses are (asymptotic) standard errors.
Figures in square brackets are absolute (asymptotic) "t-ratios".