Research Paper Series

Analytical Studies Branch

Divergent Inequalities – Theory and Empirical Results

by Michael C. Wolfson

(*Revised*) **No. 66**

Statistics Canada Statistique Canada

Canada

ANALYTICAL STUDIES BRANCH
RESEARCH PAPER SERIES

The Analytical Studies Branch Research Paper Series provides for the circulation, on a pre-publication basis, of research conducted by Branch staff, visiting Fellows and academic associates. The Research Paper Series is intended to stimulate discussion on a variety of topics including labour, business firm dynamics, pensions, agriculture, mortality, language, immigration, statistical computing and simulation. Readers of the series are encouraged to contact the authors with comments, criticisms and suggestions. A list of titles appears inside the back cover of this paper.

Papers in the series are distributed to Statistics Canada Regional Offices, provincial statistical focal points, research institutes, and specialty libraries. These papers can be downloaded from the internet at *www.statcan.ca*.

To obtain a collection of abstracts of the papers in the series and/or copies of individual papers (in French or English), please contact:

Publications Review Committee
Analytical Studies Branch, Statistics Canada
24th Floor, R.H. Coats Building
Ottawa, Ontario, K1A 0T6
(613) 951-6325

Divergent Inequalities – Theory and Empirical Results

by Michael C. Wolfson

No. 66
(Revised)

11F0019MPE No.66
ISSN:1200-5223
ISBN: 0-662-21717-9

Price: $5.00 per issue, $25.00 annually

24 R.H. Coats Building, Ottawa, K1A 0T6
*Statistics Canada and Canadian Institute
for Advanced Research(613) 951-8216
Facsimile Number: (613) 951-5643

Revised Version July, 1997
Replaces previously released version of May 1995

I am greatly indebted to Tony Atkinson for suggesting a collaboration with James Foster to probe more deeply the question of measuring polarization, to James Foster for our joint work in developing the measurement concepts, to Milorad Kovacevic, Brian Murphy, and Geoff Rowe for valuable discussion and support on the empirical and statistical work, and to an anonymous referee for helpful comments. I remain solely responsible for any errors or omissions, and for the views expressed. A substantially shortened version of this paper is Wolfson (1994).

This paper represents the views of the authors and does not necessarily reflect the opinions of Statistics Canada.

Aussi disponible en français

Table of Contents

Abstract

Widely used summary measures of inequality or the "disappearing middle class" are potentially misleading. Divergences between evidence cited and conclusions drawn include failing to distinguish the concepts of inequality and polarization, and using scalar “inequality” measures which are not consistent with rankings based on Lorenz curves. In addition, inappropriate claims about trends in inequality can arise from focusing on only a sub-population such as full-time male workers, and failing to account for sampling variability. These divergences are illustrated using Canadian data on labour incomes over the 1967 to 1994 period.

Keywords: income inequality, income polarization, disappearing middle class, sampling error

Introduction

This paper is principally about methods of income distribution analysis, particularly the foundations for claims about the extent or trend in inequality. There has been a major increase in analysis of income distribution trends. However, this has been accompanied by a somewhat undisciplined expansion in statistical methods. The result occasionally is prose conclusions which are not supported by the statistics cited. The sources of such divergences between evidence cited and conclusions drawn are the focus of this paper.

One major divergence derives from the fundamental meaning attached to the notion of inequality in the distribution of income. Further divergences between conclusions and evidence concern the particular statistical measures, populations and income definitions used to capture the intended concept. A final concern is the margin of error in the measures commonly used to support claims of trends in income inequality. We shall consider these points in turn.

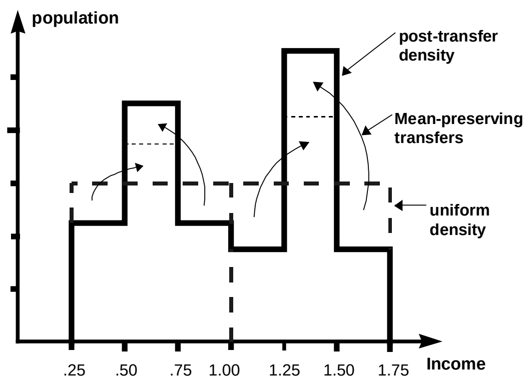
First Divergence -- Fundamental Concepts

Since the early 1980s, analysis of income distribution trends have increasingly included discussions of the "disappearing middle class" (e.g. Kuttner, 1983; Thurow, 1984). This relatively new concept is typically equated with the concept of increased income inequality. However, equating these two notions raises a fundamental conceptual issue. Levy and Murnane (1992), in their recent survey of trends in U.S. earnings inequality indicate the problem: "... *a polarization of the earnings distribution means a decline in middle class jobs"* (p1338), and later, *"Despite the variety of scalar (inequality) measures, none seems well suited to the proposition of a vanishing middle class. That proposition refers to a polarization in which observations move from the middle of the distribution to both tails. Standard inequality measures cannot distinguish this polarization from other kinds of inequality"* (p1339). However, they later illustrate the lack of clarity in this area with the inconsistent statement that, *"If the middle of the male earnings distribution was hollowed out, that fact would be registered by scalar inequality measures."* (p1351).

The kinds of phrases widely associated with discussions of polarization and a disappearing middle class provide the basis for conceptual clarification. They include "a hollowed out middle", and "individuals moving out from the middle to the tails of the distribution". These phrases imply that a more polarized income distribution is one that is more *spread out* from the middle, so there are fewer individuals or families with middle level incomes. In addition, there is a sense that this spreading out is also associated with a tendency toward *bimodality*, a clumping of formerly middle level incomes at either higher or lower levels. We take this pair of notions as central to the underlying concept of polarization.

The basic theoretical observation is that polarization understood this way is not the same concept as inequality as it has been formally defined in the literature (e.g. Atkinson, 1970). Figure 1 from Wolfson (1989) makes this clear. This graph shows two hypothetical income distribution density functions. The first is a uniform or rectangular density over the interval 0.25 to 1.75, shown by a dashed line. The second density, shown by a solid line, is clearly bi-modal, and has a somewhat depleted middle. According to our interpretation of the widely intended underlying concept of polarization or disappearing middle, this latter density is the more polarized.

Figure 1 -- Polarization and Inequality



However, the second bi-modal density has also been constructed such that according to *any* inequality measure that is consistent with the Lorenz criterion -- the "gold standard" for the concept of inequality, it is more equal. In other words, the bi-modal density has a Lorenz curve that is closer to the 45 degree line than the Lorenz curve for the uniform density. The formal proof follows simply from the fact that the bi-modal distribution can be "derived" from the uniform distribution (in several ways, one of which is) by two sets of progressive mean-preserving redistributive transfers in Atkinson's (1970) sense, as indicated by the arrows in Figure 1.

One set of equalizing income transfers is from some of the individuals in the 0.75 to 1.00 part of the income range (let's call them P's) to an identical number of individuals in the lowest part, 0.25 to 0.50 (the Q's). The P's give the Q's portions of their incomes equal to half the difference between their incomes -- 0.25 on average, so both the P's and the Q's move to the 0.50 to 0.75 income range in the bi-modal distribution. Similarly, a subset of individuals in the highest part of the income distribution with incomes between 1.50 and 1.75 (the M's say), give an average of 0.25 of their income to an equally sized set of individuals in the upper-middle part of the distribution (the N's), with incomes from 1.00 to 1.25. As a result of this set of progressive transfers, the M's and N's both end up in the same 1.25 to 1.50 income range of the bi-modal distribution. Thus, by construction, the bi-modal distribution is at the same time more polarized and more equal than the uniform distribution from which it was derived. Polarization and inequality are therefore demonstrably different concepts, as first pointed out in Love and Wolfson (1976), and reiterated in OECD (1993)[1].

This result leaves open the question of what statistics should be used to measure polarization. In the literature on the disappearing middle, in addition to inequality measures, some authors have used quintile income shares, while others have used the fraction of the population in various income ranges defined in terms of the mean or median income -- such as the proportion of the population with incomes within 25% of the median. In fact, Figure 1 has been constructed in a particularly nasty way for these kinds of statistics.

Since the distribution is symmetric, the mean is equal to the median which is 1.0. It can be shown that the *income* share of the middle third of the bi-modal distribution is lower than the income share of the middle third of the uniform distribution, while the income share of the middle two-thirds rises in the transition to the bi-modal distribution. Thus, the income shares of various middle quantile groups are *not* necessarily consistent with any meaningful formalization of the concept of polarization or disappearing middle. In turn, this means that those papers purporting to analyze the disappearance of the middle class which have used inequality indicators such as quintile shares (e.g. Levy, 1987; Beach, 1988) are unable to detect the phenomenon they claim to be studying.

Moreover, the share of the *population* with "middle level incomes" is similarly perverse in this example, going up or down depending on how "middle" is defined. This is easily seen by inspecting Figure 1. The population with incomes within 25% of the mean = median clearly falls, but the population with incomes within 50% of the mean = median rises. Thus, statistics that count the share of the population with "near middle" incomes are also *not* necessarily consistent with a sensible definition of polarization. For example, Thurow (1984) considered the proportion of the population with incomes between 75 and 125% of the median in his analysis of the disappearing middle class, while Blackburn and Bloom (1985) in a similar analysis focused on the proportion with incomes between 60 and 225% of the median.

We are thus faced with an expanding literature purporting to analyze the phenomena of inequality and the disappearing middle class, accompanied by an incoherent variety of statistical indicators. But the most basic axiom underlying the formal theory of inequality measurement -- the Pigou-Dalton condition of transfers, in turn formally equivalent to the Lorenz curve criterion -- is inconsistent with the concept of polarization, insofar as it is identified with the basic notions of spreadoutness from the middle and bi-modality that lie at the heart of the disappearing middle class phenomenon. An obvious resolution would be to formalize the concept of polarization in a manner analogous to the theoretical development of inequality measures.

The sequence of graphs in Figure 2 provides a sketch of just such a formalization. It turns out that there is a nice duality or complementarity between polarization and inequality. The strand of development in both cases starts at the top with a cumulative density function (CDF) for the distribution of income (Graph 2.1 – actually a pair of distributions showing equivalized family disposable income for Canada and the U.S. in 1988, Canada’s being more equal and less polarized).

For inequality measures, the conceptual foundation is associated with the Lorenz curve, shown in the lower left (Graph 2.3a). Graphically, it is useful to show one intermediate step between the CDF and the Lorenz curve. This step involves exchanging the axes of the CDF so population percentiles are ranged along the horizontal axis and incomes along the vertical. The result is Jan Pen's (1973) "parade of dwarfs (and a few giants)". Graph 2.2a shows this parade after dividing each individual income by the mean income. Integrating this normalized "parade" curve moving right from the origin yields the usual Lorenz curve in Graph 2.3a.

In order to formalize the concept of polarization, we can follow a similar and parallel path of graphical transformations of the initial pair of CDFs. To begin, it is useful to note two basic approaches for measuring the size of the middle class. One starts in “income space”, and asks for a given range of incomes (“M” along the horizontal axis in Graph 2.1), how large a share of the total population has incomes in that range (“S” along the vertical). The other (dual) approach starts with a population range in “people space” (“S”), and asks how wide is the range of incomes that cover this population (“M”).

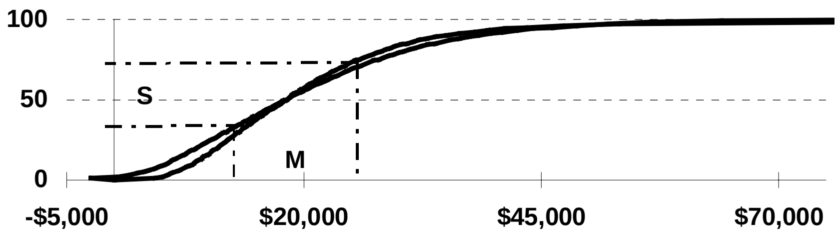
Most of the statistics being used to describe polarization (e.g. the proportion with incomes between 75 and 150% of the median) start in “income space”. However, it will be more convenient to develop the formal notion of polarization starting in “people space”. We therefore start the graphical development in the same way, by exchanging the axes of the CDFs in Graph 2.1 to form Pen’s parade of dwarfs, as in the construction of Graph 2.2a, but we then continue with the following sequence of operations:

- rather than dividing incomes by the mean, individuals' incomes along the vertical axis are normalized by dividing by the median;
- the horizontal axis is then shifted up to touch the resulting median-normalized "parade" at the mid-point of the horizontal axis, the 50th population percentile, which is (by definition) the median income (now equal to one as a result of the normalization); and
- the first half of the "parade" curve for the 50% of the population with incomes below the median (which now lies below the horizontal axis) is then flipped around the horizontal axis.

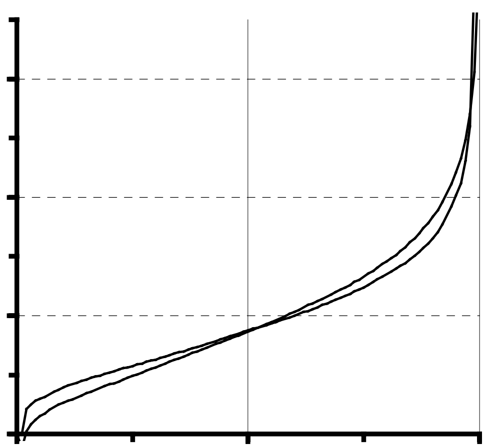
The result is the curve looking a bit like a lopsided gull shown in Graph 2.2b. It shows, for any population percentile along the horizontal axis, how far its income, expressed as a proportion of the median, is from the median. The curve therefore gives an indication of how "spread out" from the middle (50th percentile) the distribution of income is. (For any given middle range S along the horizontal axis, M is now the sum of the heights of this “spreadoutness” curve at the endpoints of the range.) A less spread out distribution (i.e. one with a larger middle class) will have a curve that is lower (and if it is everywhere lower, a Lorenz curve that is higher).

Figure 2 Inequality and Polarization, Parallel Strands of Graphical Development

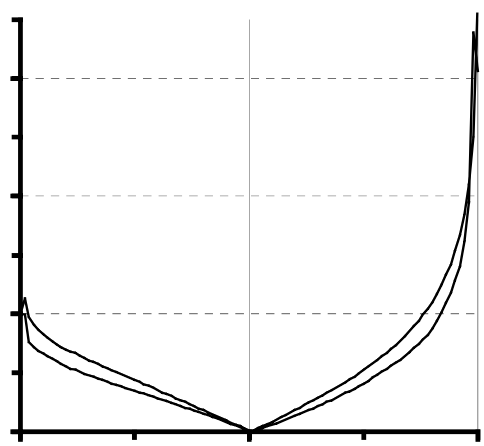
2.1 a pair of cumulative density functions



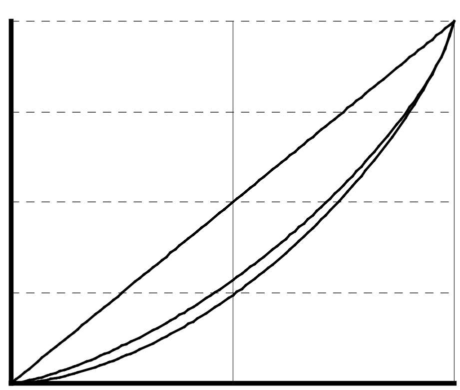
2.2a- Pen's "Parade of Dwarfs"



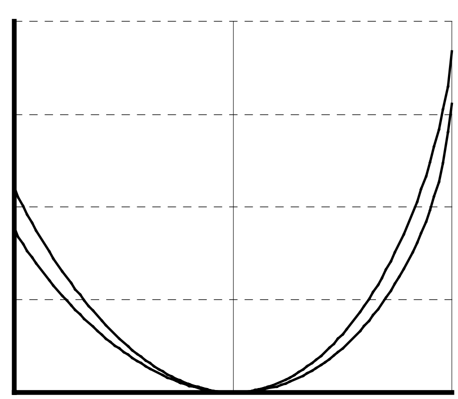
2.2b First Polarization Curve



2.3a Integrate from Left to Obtain Lorenz Curve



2.3b Integrate from Middle to Obtain Second Polarization Curve



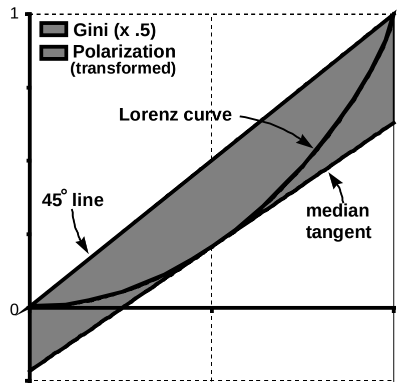
However, the concept of polarization also has a second aspect, bimodality. This is not captured by the "distance from the median" or "spreadoutness" curve in Graph 2.2b, since a progressive transfer wholly on one side of the median will result in a second curve that crosses the first. But such a transfer, like one of the pair shown in Figure 1 above, will augment the mode on its side of the median, and therefore unambiguously increase bimodality.

There is a simple transformation of the spreadoutness curve in Graph 2.2b that will make it simultaneously sensitive to both of these distributional attributes – spreadoutness from the middle and bimodality. It corresponds to the notion of moving from first to second order stochastic dominance. Formally, we integrate the "spreadoutness" curve out in both directions from the mid-point along the horizontal axis (where by construction the height of the curve is zero) to get the "cumulative spreadoutness" or polarization curve (Foster and Wolfson, 1992) in Graph 2.3b. This polarization curve not only ranks any pair of distributions in exactly the same way as the "spreadoutness" curves when they do not cross. It also ranks distributions whose spreadoutness curves cross as a result purely of increased bimodality, in exactly the way desired. This polarization curve therefore plays the same "gold standard" role for the concept of polarization as the Lorenz curve plays for inequality.

It follows that the area under this polarization curve, P, is a scalar index of polarization, just as the Gini coefficient, as (twice) the area between the 45 degree line and the Lorenz curve, is a scalar index of inequality. However, exactly analogous with Lorenz curves, it is still possible to have crossing polarization curves. Thus, polarization curves (like Lorenz curves) induce only a partial ordering over income distribution densities with respect to the sizes of their middle classes and degree of bimodality, while the area under the polarization curve P (like the Gini coefficient) induces a complete ordering.

These two strands of development, with their common starting point in the cumulative income distribution density function, can now be brought together again in a nice extension of the Lorenz curve. Figure 3 shows a typical Lorenz curve. The key addition is the tangent line to the Lorenz curve at the 50th population percentile, with the vertical axis extended down to meet this tangent. It turns out that the polarization curve just described is closely related to the Lorenz curve. If we first renormalize the vertical axis of the polarization curve (Graph 2.3b) by multiplying by the ratio of the median to the mean, and then tilt the horizontal axis until it has the same slope as the tangent to the Lorenz curve at the 50th population percentile, this transformed polarization curve is identical to the Lorenz curve!

Figure 3 -- A New Measure of Polarization Based on the Lorenz Curve



In turn, the area P under the polarization curve in Figure 2.3b, our scalar indicator of the extent of polarization or the size of the middle class, is a simple transform of the lightly shaded area in Figure 3. Specifically, the lightly shaded area in Figure 3 between the tangent line and the Lorenz curve is T - Gini / 2; and P* of Figure 2.3b is (T - Gini / 2) / mtan; where mtan = "median tangent" = m / μ = the slope of the tangent to the Lorenz curve at the 50th population percentile; m = median; μ = mean; and T = the area of the trapezoid defined by the 45 degree line and the median tangent = the vertical distance between the Lorenz curve and the 45 degree line at the 50th percentile = 0.5 - L(.5) = the difference between 50% and the income share of the bottom half of the population (which latter, L(.5), we refer to as the "median share").

P has a minimum of zero for a perfectly equal distribution of income, and a value of 0.25 for a perfectly bimodal distribution with half the population at zero income and the other half at 2 μ (with the median deemed to equal μ in this case).[2] In order to have an index with a similar range to the Gini (i.e. in the [0, 1] interval if there are no negative incomes), we shall arbitrarily define P to have four times the area discussed, so that it's formula becomes P = 2 (2 T - Gini) / mtan.

The augmented Lorenz curve in Figure 3 allows us to clarify where the conflicts between inequality and polarization arise, and why the concepts have so often been confused. If there is an "equalizing transfer" of income (in the sense of the Pigou-Dalton condition of transfers) from an individual above the median to an individual with income below the median (and the transfer is not so large that it causes either to cross the median), then both inequality and polarization decline. In terms of the diagram, such a transfer of income will move the Lorenz curve closer to the 45 degree line thereby reducing the Gini, and it will also generate a parallel upward shift in the tangent line at the 50th population percentile such that it can be shown that P also falls.[3] By virtue of this class of examples, there are clearly many situations where inequality and polarization rankings will agree.

The two concepts will disagree, however, when there are equalizing transfers entirely on one side of the median -- exactly as in Figure 1 earlier. In these cases, the median tangent curve is unaffected by the transfer, but the portion of the Lorenz curve on the affected side of the median moves closer to the 45 degree line. Such a shift in the Lorenz curve necessarily reduces the Gini coefficient, and correspondingly increases the polarization measure P. This kind of divergence between inequality and polarization could, of course, be merely a theoretical curiosum. An important question is whether in practice we may see divergent trends in the two kinds of attributes of income distributions, and illustrations are provided later.

The demonstration that inequality as formalized is not always in accord with the concept of polarization reopens the question of the axiomatic foundation of inequality measures. Specifically, it raises questions about the Pigou-Dalton condition of transfers. As noted by Amiel and Cowell (1989; fn 14), Pigou was doubtful about the validity of this axiom. Moreover, in their survey of almost one thousand undergraduate economics students (most before they had studied this topic), a majority rejected this axiom as part of their concept of inequality. At the very least, this suggests that in order to capture the concerns of the general public, summary measures based on concepts like polarization should be given equal space along with Lorenz-consistent inequality measures when describing trends in income distribution. Indeed, polarization as formalized here may be closer to the general public's vernacular concept of inequality than formal measures of inequality based on Pigou-Dalton-Lorenz-Gini concepts.

Further Divergences -- Construct Validity

In addition to a fundamental conceptual divergence between key ideas related to inequality and polarization, there are further divergences related to the notion of "construct validity". One relates to whether the specific statistical measure being used (the *indicator*) properly captures the intended underlying concept (the *indicatum*). For example, a set of questions only on whether or not the individual was working for pay, without any accompanying questions on whether he or she was actively searching for such work, would not be a valid way of determining the underlying construct or concept of labour force participation. Many purported statistical measures of income inequality in current use fail this criterion of construct validity.

Consider only the most widely agreed concept, that inequality is related to the partial ordering induced by the closeness of the Lorenz curve to the 45 degree line (notwithstanding the criticisms just made from the viewpoint of polarization). The most serious failure of construct validity is when there is inconsistency -- a more equal distribution according to the Lorenz ranking being measured as more *un*equal. Unfortunately, a number of measures are in wide use that fail on this count. The most common is the variance of logs (e.g. Karoly, 1992; Davis, 1992; Katz and Murphy, 1992). As pointed out both in Love and Wolfson (1976) and in Cowell (1977), for *equal*izing transfers above about 2.7 times the mean income (often at about the 95th percentile), the variance of logs will indicate an *in*crease in inequality.[4] Again, it is reasonable to ask whether this is merely a theoretical curiosum; empirical results presented below suggest it is not.

Another commonly used set of inequality indicators is ratios or differences in quantile income cut-points or their logs. For example, Davis (1992) focuses on the log of the ratio of a higher percentile wage to a lower percentile wage (e.g. 90th to 10th, 90th to median, and median to 10th), while the OECD (1993) and Atkinson et al (1995) analyses are based on unlogged ratios. These measures also fail the basic criterion of construct validity. A simple numerical example is a society with three individuals having incomes (1, 5, 9). If the middle individual gives one unit of income to the bottom individual, the resulting distribution (2, 4, 9) is clearly more equal according to the Lorenz curve criterion. However, the ratio of the 90th percentile income to the median increases, indicating an opposite direction of change.

This simple numerical example is admittedly extreme. It is much more likely that many of the situations where these income percentile ratios change are associated with crossing Lorenz curves. A minimal solution to this form of construct invalidity is simply not to refer to these statistics as inequality indicators or measures of inequality. They do in fact have a straightforward interpretation: the ratio of any given income percentile to the median is simply the height of Jan Pen's (1973) median- (rather than mean-) normalized "parade curve" at that population percentile (Figure 2.2a).

A more appropriate solution is to use a carefully chosen set of these kinds of indicators, rather than one or two in isolation. For example, Juhn et al. (1993) complement their inequality trend analysis based on the variance of logs with graphs showing changes in log wages for single percentiles from the 10th to the 90th. Since they generally find proportionate wage changes that are monotone increasing with percentile, such changes are in fact consistent with the Lorenz "gold standard" criterion for inequality. (However, their exclusion of the bottom and top deciles, means that they have omitted from analysis the top decile which is most likely the part of the distribution where any Lorenz inconsistency of the variance of logs will show up.)

Given that many purported inequality measures in wide use are invalid (in the sense of construct validity), what statistics should be used? It has generally been impractical to produce myriad graphs of Lorenz curves on tracing paper (or a computer screen) and compare them visually. One option is a carefully chosen small set of inequality and polarization measures.

For Lorenz inequality (i.e. indicators consistent with the partial ordering induced by Lorenz curves), a reasonable choice, to use Cowell's (1977) terminology, is one each of a bottom-sensitive, middle-sensitive, and top-sensitive inequality measure, each strictly consistent with Lorenz curve orderings.[5] If all three measures agree in a comparison of two income distributions, we can then be moderately sure that their Lorenz curves do not cross (at least to any substantial extent). However, if they do disagree, we know the Lorenz curves do cross, and hence that no unambiguous ranking is possible.

In an earlier analysis (Wolfson, 1986), we suggested that a preferred set of bottom-, middle- and top-sensitive inequality measures is the exponential measure (Exp), Gini, and (squared) coefficient of variation (CV) respectively. The only unfamiliar measure may be the exponential.[6] It was introduced because it has the advantage over other bottom-sensitive measures like the Theil-Entropy, Theil-Bernouilli (also referred to as the Mean Logarithmic Deviation), and members of the Atkinson (1970) family that it does not explode with zero or near-zero incomes. In effect, this set of three inequality measures gives priority to construct validity, to being well-defined for all conceivable levels of income (i.e. including zero and even negative amounts), and to having a high likelihood of indicating crossing Lorenz curves and therefore ambiguous Lorenz orderings. Unfortunately, these priorities rule out most inequality measures that are “nicely” decomposable (Shorrocks, 1980).

However, as illustrated empirically below, this carefully chosen set does not provide very good detection of crossing Lorenz curves. Thus, in a different sense of construct validity, taking our fundamental concept of inequality to be the *parital* ordering over distributions induced by Lorenz curves, the concern is whether this parsimonious set of valid (in the first sense of construct validity) inequality measures reliably indicates both unambiguous and ambiguous inequality rankings.

Turning to the concept of polarization, the obvious candidate for a valid (in the first sense) measure is P as defined above. A much simpler and more convenient set of measures is the proportion of the population with incomes, say, between 75% and 150% of the median, as well as a number of other similar ranges (e.g. 60 to 225%). However, as noted in connection with the discussion of Figure 1, such individual measures are *not* necessarily consistent with the formal concept of polarization that has been developed, namely the *parital* ordering induced by polarization curves of the kind shown in Figure 2.3b. They are therefore *in*valid measures of polarization in the first sense.

Moreover, crossing polarization curves are possible, just as are crossing Lorenz curves. Thus the complete ordering induced by our P statistic may disguise crossing polarization curves. Also, despite the very close relationship between the Lorenz curve and the polarization curve shown in Figure 3, crossing Lorenz curves do *not* necessarily imply similarly crossing polarization curves (or vice versa). Intuitively from Figure 3, recall the polarization curve depends not only on the "curvature" of the Lorenz curve, but also on the slope and height of the tangent to the Lorenz curve at the median. More formally, if we define the polarization curve at population percentile p as P(p) and the Lorenz curve as L(p), then $P(p) = (\mu / m) [L(p) - L(.5)] + (.5 - p)$. The clear implication is that the "gold standard" for polarization rankings should be inspection of the polarization curves P(p).

One final point concerns ratios like the 90th percentile income to the median or 10th percentile income, for example as highlighted in Atkinson et al (1995). As already discussed, these are *not* necessarily consistent -- and can indeed be inconsistent -- with the Lorenz curve-based partial ordering. It can also be shown that they are not necessarily consistent with the polarization curve-based partial ordering either. These statistics therefore have no redeeming features for the measurement of formal concepts of inequality or polarization, particularly when used in isolation without any valid measures. They simply describe a few ordinates on a median-normalized version of Pen's "Parade of Dwarfs" curve (Figure 2.2a). The only factors that can account for their continuing use is their understandability, and their wide availability (the reason they were used in OECD, 1993).

Further Divergences -- Statistical Problems

A further set of divergent inequalities may arise for statistical reasons. One concern is sampling variability. The vast majority of analyses of trends in income inequality completely omit any consideration of the underlying sampling errors. Notable exceptions include Burkhauser et. al. (1996), Bishop et. al. (1993) and Karoly (1992), though in none of these cases is account taken of the complexity of the underlying samples. For example, the U.S. March Current Population Survey and the Canadian Survey of Consumer Finance both have complex multi-stage cluster designs. Love and Wolfson (1976), using a method of half sample replication, estimated the variance of the Gini coefficient, and found that the complex sample design resulted in variances for family income almost twice what one would have expected if the data had come from a simple random sample.

More recently, Kovacevic and Binder (1997) have used an estimating equation approach at the level of sample clusters to estimate variances of several inequality and polarization measures as well as the ordinates of Lorenz and polarization curves, taking full account of the complex sample design. Their analysis was based on the same 1991 wage distribution data from the Canadian Survey of Consumer Finance used in this analysis (sample sizes on the order of 50,000 individuals). Essentially, the implied 95% confidence intervals for summary distributional statistics like the Gini coefficient and polarization P suggest as a general guideline that only the first two digits of any of the widely used inequality measures have any statistical reliability; for a top-sensitive measure like the squared coefficient of variation, only the first digit has any statistical reliability.

There are many other statistical questions that can influence results.[7] Given current interest in the distribution of wages, one important question is the population chosen for analysis. The general consensus in studies like those of Davis (1992) and Karoly (1992) is that wage inequality has been increasing over the past decade or more. However, the populations included in the analyses range from all individuals with positive labour income (both employment and self-employment) to only full-time, full-year male workers.[8] Moreover in Davis (1992), because of the limited international data available, the comparisons among countries may be contaminated by the quite different populations covered in each country's data. Thus, to carry on the theme of this paper, inequalities may diverge for the simple reason that like populations are not being compared with like.

Empirical Results

Five potential sources of divergence in inequality results have been noted:

- conceptual differences between Lorenz inequality and polarization;
- small sets of scalar measures failing to indicate ambiguities in partial orderings (i.e. crossing Lorenz or polarization curves);
- construct validity problems for purported inequality measures which are not Lorenz consistent;
- differing populations of interest; and
- sampling variability.

In this section, empirical results from a time series of Canadian Surveys of Consumer Finance are used to illustrate and assess these divergences.

Two sets of time series of distributional statistics are examined -- one for the distribution of labour income for full-time male workers, and one for the distribution of labour income for all individuals with annual labour income of at least 5% of the average wage. These two populations are denoted "FT Males" and "All ELFPs" (ELFP = effective labour force participant) respectively. Labour income includes wages and salaries, military pay and allowances, and self-employment income (which may be negative). FT Males were age 18 to 64, worked at least 48 weeks, and indicated that they mostly worked full-time. All ELFPs include females, were also age 18 to 64, but had no other restrictions on their weekly hours or annual weeks of work.

Data for these two populations and for selected years are shown in Table 1.[9] Two distinct sets of statistics are given -- the first indicating inequality, and the second polarization. The first three of the inequality indicators have been our preferred set of valid inequality measures -- the top-sensitive (squared) coefficient of variation (CV), the middle-sensitive Gini coefficient, and the bottom-sensitive Exponential measure. These are augmented by two further valid bottom-sensitive measures, the Theil-Entropy and the Theil-Bernouilli measures.[10] Finally, the last three are *invalid* inequality measures:

- the variance of logarithms;
- the ratio [ln (D9/median) / ln (median / D1)], denoted “SJDavis” since it was highlighted in the inequality analysis in Davis (1992); and
- the ratio D9 / D1 as used in OECD (1993) and Atkinson et al (1995)

where D9 and D1 represent the 90th and 10th percentile incomes respectively. These measures have been included in order to illustrate their construct *in*validity in practice.

Table 1 Selected Inequality and Polarization Indicators

	1967	1973	1981	1986	1988	1990	1991	1992	1993	1994
All ELFPs										
Valid Inequality										
Squared CV (.0429)*	0.577	0.605	0.542	0.617	0.628	0.642	0.664	0.660	0.670	0.643
Gini (.0027)*	0.363	0.378	0.377	0.396	0.395	0.398	0.402	0.404	0.405	0.399
Exponential (.0012)*	0.446	0.451	0.450	0.458	0.457	0.458	0.460	0.461	0.462	0.459
Other Inequality										
Theil-Entropy	0.243	0.259	0.251	0.276	0.277	0.280	0.288	0.290	0.292	0.283
Theil-Bernouilli	0.288	0.309	0.311	0.340	0.337	0.339	0.347	0.357	0.357	0.346
Invalid Inequality										
Variance of Logs	0.699	0.746	0.770	0.830	0.818	0.818	0.839	0.873	0.869	0.847
S.J.Davis	0.471	0.517	0.502	0.523	0.528	0.549	0.536	0.515	0.541	0.517
90th / 10th pct'ile	8.220	9.090	9.710	10.450	10.410	10.280	10.540	11.150	11.120	10.660
Polarization										
population share (%) in income range										
75-150% median	41	37	36	32	32	34	32	32	32	33
60-225% median	67	64	62	59	59	58	58	58	58	59
range of income / median covering middle										
40-60% population	0.341	0.400	0.396	0.434	0.424	0.419	0.437	0.440	0.453	0.447
30-70% population	0.700	0.792	0.815	0.912	0.900	0.877	0.889	0.926	0.935	0.886
20-80% population	1.123	1.255	1.311	1.403	1.393	1.388	1.410	1.412	1.449	1.385
median share	0.243	0.229	0.226	0.213	0.215	0.214	0.210	0.208	0.207	0.211
median / mean	0.900	0.871	0.883	0.861	0.856	0.853	0.852	0.859	0.843	0.863
Polarization (P, .0042)*	0.337	0.376	0.384	0.415	0.410	0.409	0.417	0.420	0.429	0.413
* 1991 standard error										
FT Males										
Inequality										
Squared CV	0.328	0.273	0.234	0.288	0.306	0.327	0.331	0.332	0.355	0.337
Gini	0.261	0.245	0.242	0.263	0.266	0.274	0.272	0.273	0.275	0.274
Exponential	0.412	0.407	0.405	0.412	0.413	0.415	0.414	0.415	0.416	0.415
Other Inequality										
Theil-Entropy	0.135	0.116	0.109	0.130	0.133	0.141	0.139	0.141	0.141	0.145
Theil-Bernouilli	0.140	0.117	0.120	0.143	0.142	0.151	0.145	0.149	0.149	0.149
Invalid Inequality										
Variance of Logs	0.308	0.248	0.273	0.331	0.316	0.339	0.312	0.328	0.328	0.318
S.J.Davis	0.774	0.827	0.775	0.654	0.701	0.744	0.726	0.688	0.713	0.717
90th / 10th pct'ile	3.370	3.100	3.250	3.680	3.590	3.690	3.780	3.760	3.690	3.710
Polarization										
population share (%) in income range										
75-150% median	60	60	58	56	55	53	54	54	54	54
60-225% median	81	84	83	79	79	79	78	78	78	78
range of income / median covering middle										
40-60% population	0.190	0.194	0.201	0.220	0.224	0.249	0.237	0.230	0.230	0.246
30-70% population	0.415	0.418	0.430	0.474	0.484	0.496	0.498	0.500	0.500	0.504
20-80% population	0.705	0.702	0.712	0.763	0.787	0.803	0.809	0.799	0.799	0.831
median share	0.319	0.328	0.327	0.313	0.311	0.306	0.307	0.307	0.307	0.305
median / mean	0.917	0.920	0.936	0.945	0.924	0.913	0.912	0.923	0.923	0.919
Polarization (P)	0.219	0.216	0.221	0.236	0.241	0.248	0.250	0.246	0.246	0.251

The second set of statistics is related to the concept of polarization. The first five statistics count the proportion of the population with "middle class" labour incomes, though from two different perspectives. The first pair, denoted "population share (%) in income range (of) 75-150% or 60-225% median", give the proportions of individuals with incomes between 75 and 150 percent of the median, and those with incomes between 60 and 225 percent of the median, respectively. These statistics measure the size of the middle class defined in terms of a range of median-normalized incomes. The next three statistics effectively exchange the axes by defining the middle class in "people space" rather than "income space". These statistics are based on symmetric percentile ranges of the population -- within 10, 20, and 30 percent of the 50th percentile, denoted "40-60% population", "30-70% population", and "20-80% population" respectively. For each of these "people space" ranges, the corresponding range of incomes they span, divided by the median, is the statistic given. Thus, for example, if the figure for "40-60% population" is 0.341 (as shown for All ELFPs in 1967), this is the 60th percentile income minus the 40th percentile income divided by the median. Even though Figure 1 above shows that any one of these statistics may be misleading by itself, agreement amongst a set is more likely to indicate an unambiguous change in polarization as we have formalized the concept.

The last three polarization-related statistics are all derived from the polarization/Lorenz curve shown in Figure 3. The first two are simply ingredients in the calculation of the summary polarization index P, though they are interesting indicators in their own right. One is the "median share" mentioned earlier -- the share of income accruing to the bottom half of the population. This in turn is exactly the height of the Lorenz curve halfway along the horizontal axis, i.e. at the 50th percentile (hence more properly considered an inequality indicator). Also, (0.5 - median share) is the vertical distance between the 45 degree line and the Lorenz curve at the 50th percentile, hence the area T of the trapezoid enclosing the Lorenz curve in Figure 3. The second statistic is the ratio of the median to the mean income, m / μ. In addition to the graphical interpretation of this being the slope of the tangent to the Lorenz curve at the 50th percentile (mtan), this ratio is also an indicator of the skewness of the distribution. Finally, the last statistic is the polarization measure P defined above. Higher P means more polarization, and a smaller middle class.

We turn now to an examination of Table 1 to explore the varieties of divergent inequality. For the time being, we ignore sampling variability and assume the distributions have been observed with infinite precision.

The first kind of divergence is between Lorenz inequality and polarization. Generally, measures indicating the two concepts move in the same direction for pairwise comparisons of adjacent income distributions. But from 1973 to 1981 for both labour force definitions, all of the valid Lorenz consistent inequality measures decline except the Theil-Bernouilli. At the same time, all but one of the polarization measures indicate an increase – the shares of the two population groups with middle earnings (75 - 150 and 60 - 225) all decline; the widths of all but one of the income intervals required to enclose various symmetric middle populations all increase, as does the summary polarization index. Thus, according to what is by far the most common analytical style in the literature, where reference is made only to a handful of summary measures rather than to the full distribution, we have an instance where polarization and inequality move in opposite directions.

However, this is also an instance of the second kind of divergence, where even our carefully chosen threesome of valid inequality measures fails to indicate crossing Lorenz curves. A sequence of Lorenz curve ordinates is shown in Table 2, along with a

Table 2 **Lorenz Curve Ordinates (%) at Selected Population Percentiles, All ELFPs, Canada, 1967 - 1994**

Population Percentile	s.e. (%)	1967		1973		1981		1986		1988		1990		1991		1992		1993		1994
5	0.005	0.5	=	0.5	=	0.5	+	0.4	=	0.4	=	0.4	=	0.4	=	0.4	=	0.4	=	0.4
10	0.016	1.3	=	1.3	+	1.2	=	1.2	=	1.2	=	1.2	=	1.2	+	1.1	=	1.1	-	1.2
15	0.028	2.6	+	2.5	+	2.4	+	2.3	=	2.3	=	2.3	=	2.3	+	2.1	-	2.2	=	2.2
20	0.043	4.4	+	4.1	+	4.0	+	3.7	-	3.8	=	3.8	+	3.7	+	3.5	=	3.5	-	3.7
25	0.059	6.7	+	6.2	+	6.0	+	5.5	=	5.6	=	5.6	=	5.5	+	5.3	=	5.3	-	5.5
30	0.078	9.4	+	8.7	+	8.5	+	7.8	=	7.9	=	7.9	+	7.7	+	7.5	=	7.5	-	7.7
35	0.096	12.5	+	11.7	+	11.4	+	10.5	=	10.6	=	10.6	+	10.4	+	10.1	=	10.1	-	10.4
40	0.115	16.1	+	15.0	+	14.7	+	13.6	=	13.8	=	13.8	+	13.4	=	13.2	=	13.2	-	13.5
45	0.135	20.0	+	18.8	+	18.5	+	17.2	=	17.4	=	17.4	+	17.0	+	16.7	=	16.7	-	17.1
50	0.153	24.3	+	22.9	=	22.6	+	21.3	=	21.5	=	21.4	+	21.0	=	20.8	=	20.7	-	21.1
55	0.173	29.0	+	27.5	=	27.3	+	25.8	=	25.9	=	25.9	+	25.5	=	25.3	=	25.2	-	25.7
60	0.189	34.1	+	32.5	=	32.3	+	30.8	=	30.9	=	30.8	=	30.5	=	30.2	=	30.1	-	30.7
65	0.208	39.5	+	38.0	=	37.9	+	36.3	=	36.3	=	36.3	=	35.9	=	35.7	=	35.6	-	36.2
70	0.222	45.4	+	43.9	=	44.0	+	42.4	=	42.4	=	42.2	=	41.8	=	41.7	=	41.6	-	42.2
75	0.239	51.7	+	50.4	=	50.5	+	49.1	=	49.0	=	48.7	=	48.4	=	48.4	=	48.2	-	48.8
80	0.256	58.5	+	57.4	=	57.8	+	56.4	=	56.2	=	55.9	=	55.6	=	55.7	=	55.6	=	56.0
85	0.269	65.9	+	65.1	-	65.7	+	64.5	=	64.3	=	63.9	=	63.7	=	63.8	=	63.7	=	64.1
90	0.281	74.1	=	73.7	-	74.6	+	73.6	=	73.3	=	72.9	=	72.8	=	72.9	=	72.9	=	73.2
95	0.286	83.8	=	83.8	-	84.9	+	84.1	=	83.8	=	83.5	=	83.4	=	83.5	=	83.5	=	83.7
97	0.280	88.5	=	88.5	-	89.7	+	88.9	=	88.7	=	88.5	=	88.4	=	88.5	=	88.5	=	88.6
99	0.242	94.4	=	94.5	-	95.4	+	94.9	=	94.7	=	94.7	=	94.5	=	94.6	=	94.4	=	94.6

set of symbols indicating statistically significant differences, which are discussed later. Similarly, Table 3 shows a sequence of corresponding polarization curve ordinates.

The fact that one of the valid inequality measures shown (the most bottom-sensitive, though not one of our threesome), the Theil-Bernouilli, moves differently from the other valid inequality measures in Table 1 correctly signals crossing Lorenz curves from 1973 to 1981, as shown in Table 2 for All ELFPs. (Recall that we are ignoring sampling variability for the time being.) In most cases, though, the threesome of indicators correctly indicates whether or not the underlying Lorenz curves cross. For example the threesome moves inconsistently from 1986 to 1988 and from 1991 to 1992 when the Lorenz curves also cross (still ignoring sampling variability).

The general agreement amongst the polarization indicators (all but one of which are *in*valid in the strict sense) in ranking the 1973 All ELFP distribution as more polarized than that for 1981 correctly indicates "polarization dominance" in terms of the underlying polarization curves shown in Table 3 (though we do see a conflict between the "40 - 60% population" indicator and the "gold standard" unambiguous ranking of the polarization curves).

Table 3 **Polarization Curve Ordinates (%) at Selected Population Percentiles, All ELFPs, Canada, 1967 - 1994**

Population Percentile	s.e. (%)	Year 1967		1973		1981		1986		1988		1990		1991		1992		1993		1994
5	0.142	18.5	-	19.3	-	20.0	-	20.7	+	20.3	=	20.4	-	20.8	-	21.3	+	20.9	=	21.0
10	0.134	14.4	-	15.2	-	15.8	-	16.6	+	16.3	=	16.3	-	16.8	-	17.1	+	16.8	=	16.9
15	0.122	10.9	-	11.6	-	12.1	-	12.9	+	12.6	=	12.6	-	13.1	=	13.2	=	13.1	=	13.1
20	0.106	7.9	-	8.4	-	8.9	-	9.6	+	9.3	=	9.4	-	9.7	=	9.9	+	9.6	-	9.8
25	0.088	5.4	-	5.8	-	6.2	-	6.6	+	6.4	=	6.5	-	6.8	=	7.0	+	6.7	-	6.9
30	0.067	3.4	-	3.7	-	4.0	-	4.3	+	4.1	=	4.2	-	4.4	=	4.5	+	4.3	-	4.5
35	0.047	1.9	-	2.1	-	2.3	-	2.5	+	2.3	=	2.3	-	2.6	=	2.5	+	2.4	-	2.6
40	0.029	0.9	=	0.9	-	1.1	=	1.1	=	1.0	-	1.1	=	1.1	-	1.2	=	1.1	-	1.2
45	0.011	0.2	-	0.3	-	0.4	+	0.2	+	0.2	-	0.3	=	0.3	+	0.2	-	0.3	-	0.4
50	0	0.0	=	0.0	=	0.0	=	0.0	=	0.0	=	0.0	=	0.0	=	0.0	=	0.0	=	0.0
55	0.023	0.2	-	0.3	=	0.3	+	0.2	+	0.1	-	0.3	=	0.3	=	0.2	-	0.3	=	0.3
60	0.046	0.9	-	1.0	=	1.0	=	1.0	=	1.0	=	1.0	-	1.1	+	0.9	-	1.2	=	1.1
65	0.093	1.9	-	2.3	=	2.3	=	2.4	=	2.3	=	2.5	=	2.5	=	2.3	-	2.7	=	2.5
70	0.129	3.5	-	4.1	=	4.2	-	4.5	=	4.4	=	4.4	=	4.4	=	4.3	-	4.8	+	4.4
75	0.188	5.5	-	6.6	=	6.6	-	7.3	=	7.1	=	7.0	=	7.2	=	7.1	-	7.6	+	7.1
80	0.259	8.0	-	9.6	=	9.9	-	10.8	=	10.5	=	10.5	=	10.6	=	10.6	-	11.4	+	10.4
85	0.334	11.2	-	13.4	=	13.8	-	15.2	=	15.0	=	14.9	=	15.1	=	15.1	-	16.0	+	14.8
90	0.420	15.4	-	18.3	=	18.9	-	20.8	=	20.5	=	20.4	=	20.8	=	20.6	-	21.9	+	20.4
95	0.538	21.1	-	24.9	=	25.6	-	28.0	=	27.8	=	27.8	=	28.2	=	28.0	-	29.5	+	27.5
97	0.595	24.4	-	28.3	=	29.0	-	31.5	=	31.5	=	31.7	=	32.1	=	31.8	-	33.4	+	31.2
99	0.683	28.9	-	33.2	=	33.5	-	36.5	=	36.5	=	37.0	=	37.3	=	36.9	-	38.4	+	36.1

Thus, our example of the first kind of divergence, between inequality and polarization, turns out to be more apparent than real. It is observed when comparing the valid threesome of inequality measures to our suite of polarization indicators, but in this case the three valid summary scalar measures of inequality failed to indicate a crossing of the underlying Lorenz curves.

The third kind of divergence concerns the construct validity of individual measures. For the variance of logarithms, there are at least two instances of apparent divergences in Table 1. For All ELFPs, the variance of logarithms moves in the opposite direction to all the Lorenz-consistent inequality measures comparing period 1991 to 1994. In this case, Table 2 shows there is a slight crossing of Lorenz curves – the 1991 Lorenz curve is slightly higher around the 15th percentile, while the 1994 Lorenz curve is everywhere else at the same level or higher. A similarly apparent divergence for the variance of logarithms is shown for the FT Males population from 1986 to 1991; though again, the underlying Lorenz curves cross somewhat at the bottom quintile (not shown).

The other two *invalid* inequality indicators can also give false results. For example, the D9 / D1 ratio decreases from 1988 to 1990 for all ELFPs, while all the valid inequality measures increase, and. as shown in Table 2, the Lorenz curves do not cross., so the inequality ranking is unambiguous. Similarly, the SJDavis measure declines for All ELFPs from 1990 to 1991 even though all the valid inequality measures increase, and again the Lorenz curves do not cross. Thus, these two examples show *in*validity in practice not only with respect to a set of valid scalar inequality measures, but also with the gold standard of Lorenz dominance.

A fourth kind of inequality divergence concerns the choice of population. As already noted, researchers like Davis (1992) have focused on full-time male populations. In principle, however, focusing on only a subset of the working population may neglect the impacts of contemporaneous trends, such as increasing female labour force participation, an increase in part-time work, and changes in self-employment. In general, Table 1 shows that inequality among FT Males is lower and more stable over time than among All ELFPs. The clearest divergence in trends associated with choice of population is in the 1967 to 1973 period. Inequality moves significantly in opposite directions for the two populations. Moreover, these are unambiguous rankings, as neither pair of underlying Lorenz curves cross. In addition, polarization is stable or declining over this period for FT Males, while it clearly increases for All ELFPs.

The final kind of divergent inequality is where authors interpret the data as showing trends where no trends exist -- because the changes are not statistically significant. Table 1 shows three digits for most of the statistics, while it was noted earlier that taking account of sampling error would leave at most two digits statistically significant, and often only one for top-sensitive measures like the squared CV.

More specifically Table 4, taken from Kovacevic and Binder (1997, and personal communication) gives consistent estimates of the coefficient of variation (CV) for the 1991 All ELFP data in Tables 1 to 3, taking full account of the underlying complexities of the sample design.[11] The table also shows the 95 percent confidence interval (95% C.I., assuming normality), and the design effects associated with the complex sample design (measured as the ratios of the correctly estimated variances to variances estimated as if the survey was a simple random sample). Using these estimates of sampling variability, the columns of symbols between adjacent pairs of Lorenz curve and polarization curve ordinates in Tables 2 and 3 respectively then indicate whether (+) or not (-) the Lorenz or polarization curve to the left is above the one to the right at that population percentile, or not significantly different (=) based on the simplifying assumption that the standard errors for all survey years are the same as those in 1991. (This assumption likely understates the standard errors in earlier years when sample sizes were smaller).

Table 4 – Sampling Variability for Selected Measures

Measure	Value	CV %	95% C.I.	Design Effect
CV^2	.761	5.64	0.0858	1.87
Gini	.412	0.66	0.0054	3.54
Exp	.460	0.26	0.0024	3.53
P	.399	1.07	0.0085	2.18
Lorenz (.25)	.055	1.06	0.0012	2.96
Lorenz (.50)	.210	0.72	0.0030	2.96
Lorenz (.75)	.487	0.49	0.0048	3.47
Lorenz (.90)	.728	0.38	0.0055	3.84
Polarization (.20)	.097	1.09	0.0021	3.03
Polarization (.40)	.011	2.63	0.0006	2.90
Polarization (.60)	.011	4.08	0.0009	2.82
Polarization (.80)	.115	2.25	0.0052	2.82
Polarization (.90)	.208	2.02	0.0084	3.00

The design effects in Table 4 are particularly notable. Even in the careful analysis of Bishop et. al. (1993), where an attempt was made to account for the complexities of the sample design using the method of Beach and Kaliski (1986), their design effects are about one-third those shown here (Kovacevic, personal communication reproducing the Beach-Kaliski approach on these data). As a related example, Kovacevic and Binder (1997) find a CV for 1991 median labour income of about one percent in our sample of over 50,000, while Bishop et al (1993, Table A4) show a CV for 1981 median per capita family income of 0.6%, in a sample of about 15,000. The reason they underestimate sampling variability for Canada so seriously is that they were working with a public use version of the microdata file which, to protect the confidentiality of survey respondents, does not include any information on sample design, particularly information on clustering.

If these major under-estimates of sampling variability are also relevant for the other LIS countries' data sets analysed in Bishop et al., then many of their "statistically significant" dominance relationships will become insignificant. There are similar implications for many other inequality studies.

Returning to the data being analysed here, these standard errors suggest that for All ELFPs, the only clearly statistically significant pairwise changes in Table 1 are increases over the 1967 to 1973 and 1981 to 1986 periods for the Gini and Exp measures, and P from 1967 to 1973. In turn, these significant increases in inequality over the 1967 to 1973 and 1981 to 1986 periods are supported by inspection of the underlying Lorenz curves in Table 2, which do not cross substantially.

However, the conclusion from Table 1 of no significant inequality trend over the 1973 to 1981 period is inappropriate because the underlying Lorenz curves do cross. The virtual constancy of the Gini and Exp mask offsetting and statistically significant shifts in the underlying distribution over the period. And the summary index of polarization is misleading in another sense. P shows no significant change in polarization from 1973 to 1981 and 1986 to 1988 at the 95% level, while Table 3 shows statistically significant and unambiguous changes, particularly in the bottom half of the distribution.

The clear methodological conclusion is that the only reliable indicators of trends in inequality or polarization are the full underlying Lorenz and polarization curves.

If we focus on the Lorenz curves in Table 2, the general story is statistically significant increasing inequality for half the sequence of pairwise comparisons. There was an ambiguous change from 1973 to 1981, virtual constancy over the 1986 to 1988, 1988 to 1990, and 1992 to 1993 periods, and a decline from 1993 to 1994. More broadly, inequality generally increased over the two decades from 1967 to 1986, but statistically was unchanged from 1986 to 1994.

The story with regard to polarization is generally similar. Polarization increased continually from 1967 to 1986 (with the exception of a slight crossing of polarization curves from 1981 to 1986). However, from 1986 to 1994, polarization remained generally stable. In particular, comparing the polarization curve ordinates for the (non-adjacent) years 1986 and 1994 generally shows no statistically significant differences (exceptions being the 45th and 55th percentiles).

Tables 2 and 3 together show that both Lorenz curves and polarization curves can and do cross in practice, so some rankings are ambiguous. These tables also show that the two kinds of curves do *not* cross in the same ways or at the same places, notwithstanding their close relationship.

Finally, this experience of analysing observed trends in summary measures of inequality and polarization has been somewhat frustrating. Often, changes in summary measures are not statistically significant. But sometimes apparent stability among a threesome of inequality measures deliberately chosen to be sensitive to changes in inequality throughout the income spectrum can mask substantial changes in the underlying distribution of labour income, though these are changes that involve crossing Lorenz curves.

The main argument for the use of summary measures in the first place has been their convenience compared to the tedium of actually comparing the full Lorenz curves -- even though this is the "gold standard" for judging changes in inequality. This tedium argument loses force, however, with current computing power, and the widespread availability of microdata. The spreadsheet used to generate all the tables and figures above is not that complex, and is well within the capacity of most PCs in use today. (A copy is available by contacting *wolfson@statcan.ca*).

Concluding Comments

Widely used summary statistical indicators of inequality or the "disappearing middle class" are potentially misleading. First, the fundamental concepts of inequality and polarization are distinct and do not always rank distributions the same way. Second, some measures like the variance of logarithms and the D9 / D1 ratio do not measure what most people think -- they are invalid for *inequality* analysis. Moreover, even valid inequality measures may give misleading indications compared to the “gold standard” of Lorenz dominance. Beyond these fundamental problems of clarity of concept and construct validity, claims made about trends in inequality may be inappropriate because they fail to account for sampling variability, or they should be more clearly circumscribed when only a sub-population like full-time male workers is being considered.

For all of these cases of potential divergence between evidence cited and conclusions claimed, examples have been given to show the salience of the problems. The implications can be summarized in a handful of suggestions for rigorous income distribution analyses: (1) particularly when discussing topics like the “disappearing middle class”, include P and/or related polarization measures in the suite of statistical indicators used for analysis; (2) avoid the variance of logs and income ratios like the 90th to the 10th percentile in discussions of inequality; (3) unless there is explicit analysis suggesting greater precision, consider only two digits of any inequality statistic (and only one digit for top-sensitive measures like the coefficient of variation) to be statistically significant, or as a rough rule of thumb if the income distribution data come from a sample with a complex clustered design, assume a design effect of three; (4) try to use comprehensive and consistent populations for comparison, or at least present these results as background when focusing on sub-populations; (5) try always to examine the underlying Lorenz and polarization curves as the "gold standard" for unambiguous rankings; and (6) take advantage of modern computing power to produce more comprehensive suites of statistical indicators and new kinds of tabular or diagrammatic methods for visualization of trends.

References

Amiel, Y. and F. Cowell (1989), "Measurement of Inequality: Experimental Test by Questionnaire", Discussion Paper No. TIDI/140, October, Suntory-Toyota International Centre for Economics and Related Disciplines, London School of Economics, London.

Atkinson, A.B. (1970), "On the Measurement of Inequality", Journal of Economic Theory, Vol.-2.

Atkinson, A. B. L. Rainwater, and T. M. Smeeding (1995), Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study (LIS), OECD Social Policy Studies No. 18, Paris.

Beach, C.M. (1988), "The ’Vanishing’ Middle Class?: Evidence and Explanations", Queen’s Papers in Industrial Relations, Industrial Relations Centre, Queen’s University at Kingston.

Beach, C.M. and S.F.Kaliski (1986), “Lorenz Curve Inference with Sample Weights: an Application to the Distribution of Unemployment Experience”, Applied Statistics, 35, 439-50.

Bishop, J.A., J.P.Formby, and W.J.Smith (1993), “International Comparisons of Welfare and Poverty: Dominance Orderings for Ten Countries”, Canadian Journal of Economics, XXVI, No. 3.

Blackburn, M. and D.E. Bloom (1985), "What is Happening to the Middle Class?", American Demographics, January.

Burkhauser, R.V., A.D.Crews, M.C.Daly, and S.P.Jenkins (1996), “Where in the world is the Middle Class? A Cross-National Comparison of the Vanishing Middle Class Using Kernel Density Estimates”, ESRC Working Paper 96-8, University of Essex.

Cowell, F.A. (1977), Measuring Inequality, Oxford, Philip Allan Publishers.

Davis, S.J. (1992), "Cross-Country Patterns of Change in Relative Wages", in O.J. Blanchard and S. Fischer (Eds), NBER Macroeconomics Annual, MIT Press.

Foster, J. and M.C. Wolfson (1992), "Polarization and the Decline of the Middle Class: Canada and the U.S.", Vanderbilt University and Statistics Canada, mimeo.

Juhn, C., K.M. Murphy and B. Pierce (1993), “Wage Inequality and the Rise in Returns to Skill”, Journal of Political Economy, Vol. 101, No. 3

Karoly, L. (1992), "Changes in the Distribution of Individual Earnings in the United States: 1967-1986", Review of Economics and Statistics.

Katz, L. and K. Murphy, (1992), "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", Quarterly Journal of Economics, Vol. CVII, February, Issue 1.

Kovacevic, M.S. and D.A. Binder (1997), "Variance Estimation for Measures of Income Inequality and Polarization – The Estimating Equations Approach", forthcoming, Journal of Official Statistics.

Kuttner, B. (1983), "The Declining Middle", The Atlantic Monthly, July.

Levy, F. (1987), "Changes in the Distribution of American Family Incomes, 1947 to 1984", Science, Vol.-236, pp.-923-27, May.

Levy, F. and R.J. Murnane (1992), "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations", Journal of Economic Literature, September, Vol XXX, No 3.

Love, R. and M.C. Wolfson (1976), Income Inequality: Statistical Methodology and Canadian Illustrations, Catalogue 13-559 Occasional, Statistics Canada, Ottawa, March.

OECD (1993), Employment Outlook, Paris.

Pen, J. (1973), "A Parade of Dwarfs (and a Few Giants)", in A.B.Atkinson (Ed.), Wealth, Income and Inequality, Penguin, Middlesex.

Rowe, G. (1994), "Income Statistics from Survey Data: Effects of Respondent Rounding", mimeo, Social and Economic Studies Division, Statistics Canada, Ottawa, January.

Shorrocks, A.F. (1980), “The Class of Additively Decomposable Inequality Measures”, Econometrica, Vol. 48, No. 3

Thurow, L. (1984), "The Disappearance of the Middle Class", New York Times, February 5, sect 3, p. 2.

Wolfson, M.C. (1986), "Stasis Amid Change -- Income Inequality in Canada 1965-1983", Review of Income and Wealth, December.

Wolfson, M.C. (1989), "Inequality and Polarization: Is There a Disappearing Middle Class in Canada?", in Proceedings of the Statistics Canada Symposium on Analysis of Data in Time, October 1989, Statistics Canada, Ottawa.

Wolfson, M.C. and J. Foster (1993), "Inequality and Polarization -- Concepts and Recent Trends", paper presented to the Winter Meetings of the American Statistical Association, Ft. Lauderdale, January.

Wolfson (1994), "When Inequalities Diverge", American Economic Review, May.

Notes

[1] Kolm (1966) was also aware of this distinction, though he used different terminology. He identified the Pigou-Dalton Condition of Transfers with the word "rectifiance". He then observed that, while a sequence of "rectifiants" changes in any given distribution of income would eventually bring it to complete equality (i.e. a Lorenz curve coincident with the 45 degree line), this need not imply that there would be a monotone decrease in inequality ("isophily" - love of equality, in his terms) in the process. In particular, such a sequence of transfers would cause some pairs of individuals to move further apart along the income spectrum - essentially our notion of "spreadoutness".

[2] Note that 0.25 is not necessarily the maximum. P could exceed 0.25 if half the population had a negative average income. Similarly, when there are negative incomes, the Gini can exceed 1.0. Also, for any given median share and mtan both positive, P is minimized and approaches zero for a tri-modal distribution where one individual has a very large negative income, another has a very large positive income, and everyone else has the same income in between.

[3] The shift is parallel since mtan is unaffected by the transfer. Note that eventually, such upward parallel shifts will result in the median tangent touching the 45 degree line (either at the 0^{th} percentile if $m / \mu < 1$, or the 100^{th} percentile otherwise), thereby implying a linear Lorenz curve over at least half its length. This is equivalent to a sequence of median-crossing, mean- and median-preserving transfers resulting in one half of the distribution becoming concentrated at a single income. Thus, a reduction in P can be associated with an increase in bimodality. However, since P is based on the extent of both spreadoutness and bimodality, the reduction in spreadoutness in this case more than offsets the increase in bimodality.

[4] One factor that could account for the continued popularity of the variance of logarithms is its convenient decomposition. However, the Theil-Entropy, Theil-Bernouilli and squared Coefficient of Variation (defined below) also have "nice" decompositions and do not suffer from construct *in*validity.

[5] A slightly different formulation of sensitivity to transfers at various points in the income spectrum, with similar implications, was developed in Love and Wolfson (1976).

[6] $Exp = \Sigma\, p_i \exp(-y_i / \mu)$ where p_i is the proportion of the population in the i-th income group, y_i is the average income in that group, and μ is the overall mean income.

[7] For example, no account is taken here of Rowe's (1994) finding that respondents often report their incomes in rounded amounts (e.g. to the nearest $100 or $1,000). Rowe estimates for similar earnings data to that used here that this rounding behaviour results in response errors in addition to and of the same magnitude as the sampling errors estimated by Kovacevic and Binder (1997).

[8] There is a further problem of interaction between the use of bottom-sensitive inequality measures and the choice of population. If all strictly positive earners are included as in Karoly (1992), compared with a somewhat higher *de minimus* threshold like 5% of the average wage, measures like the Theil-Entropy and Theil-Bernouilli could show spurious changes due to fluctuations in the sub-populations with only a few dollars of earnings. This problem has been encountered using Canadian data.

[9] Each distribution was first tabulated to give 100 percentile means. All the statistics shown were then calculated from these intermediate data. This does induce some approximation (grouping) error; however except for the squared CV, these are negligible when compared to results derived directly from the raw microdata.

[10] Theil-Entropy = $\Sigma (y_i / \mu) \ln (y_i / \mu)$; and Theil-Bernouilli = $- \Sigma p_i \ln (y_i / \mu)$ where p_i is the proportion of the population in the i-th income group, y_i is the average income in that group, and μ is the overall mean income.

[11] The underlying 1991 sample contained 50,701 observations in 4,201 clusters or Primary Sampling Units, in turn drawn from 1,139 strata. Note that these statistics were calculated from the full microdata, so that comparing the values here to those in Tables 1 to 3 is an indication of the approximation error resulting from using 100 percentile means as the base for deriving all the other statistics shown.