TOWARD AN EMPIRICAL ANALYSIS OF POLARIZATION

GORDON ANDERSON

ECONOMICS DEPARTMENT

UNIVERSITY OF TORONTO

August 2001

Preliminary, not to be quoted. Comments most welcome
(send to anderson@chass.utoronto,ca)

By describing polarization states in terms of stochastic dominance conditions this paper presents a taxonomy of, and tests for, polarization both between and within population distributions. Four examples from the poverty, growth and development, wage distribution and assortative pairing literatures are used to illustrate how the techniques may be used in economic applications.

Thanks for helpful comments are due Charles Beach, to the seminar memberships at Irvine and Toronto Economics Departments, the Canadian Economics Association meetings in Montreal and the European Econometric Society Meetings in Lausanne. Thanks also to Emily Hanna, Shinji Ge, Scot Rutherford and Sandra Shao for their diligent research assistance. All errors remain my responsibility.

Introduction.

The concept of "Polarization", what the Oxford English Dictionary refers to as "..the tendency to develop in two opposite directions in space, time, serial direction ..", is gaining frequent usage in economics. Foster and Wolfson (1992) Levy and Murnane (1992), Jenkins (1996), Beach and Slotsve (1996) and Beach, Chaykovske and Slotsve (1998) each employ the idea in describing the diminution of the middle class in wage and income distributions[1], Jones (1997) and Quah (1997) apply it in studying growth convergence issues. These literatures broadly interpret polarization as the disappearance of mass at the centre of an empirical distribution or the increasing distance between and intensity of, points of modality in it as it evolves through time. However the concept need not be confined to the study of within distribution changes, it can be used in assessing the relative movements of two or more distributions as they evolve (for example polarization between ethnic groups, genders, nations etc) and thus has widespread application beyond the income, wage and growth literatures cited. Whilst within and between population polarization analysis presents quite distinct empirical problems there are many common features which can be exploited.

At the theoretical level Esteban and Ray (1994) take an axiomatic approach to defining polarization indices reflecting the emergence of many poles (the empirical literatures cited above and this paper confine themselves to two) in the context of a discrete distribution and Wolfson (1994) focusses on the distinction between polarization and inequality in developing a polarization

---

[1] Levy and Murnane (1992) in reviewing U.S. earning trends note "Inequality in the male earnings group has taken the form of polarization" and further note "Despite the variety...standard inequality measures cannot distinguish this polarization from other kinds of inequalities...". Jenkins (1996) in a study of UK income distributions observes "Polarization provides a challenge for our thinking about how we assess income distributions, since the measurement tools economists have developed focus almost exclusively on changes in income levels and dispersion". Beach and Slotsve (1996) pose the question "Why look at polarization specifically" and answer "Polarization .... can be viewed as an extreme version of the more general term "inequality"... ".

measure with a Lorenz curve interpretation. While statistical tests of within population Polarization have not been developed, the analysis of "bumps" within a distribution has a genesis in a statistical literature arguing that multimodality is more easily studied in the context of a mixture of unimodal distributions[2] (there are relatively few parametrically specified multimodal distributions) and proposing tests for spotting multiple modes or dips (Cox(1966), Good and Gaskins (1980), Silverman (1981) and Hartigan and Hartigan (1985)). Contemplating the notion of within distribution polarization in the context of changes in the sub-population distributions of a mixture greatly facilitates analysis[3], unfortunately polarization is a tendency and can take place before bumps actually emerge, rendering "bump" or "dip" seeking tests less useful in detecting changes in the nature of, or tendencies towards, multiple modes. Tests for Polarisation between populations have typically taken indirect forms reflecting trends in location and scale differences and measures of the extent to which distributions overlap (see Gastwirth, Nayak and Wang (1989), Gibbons and Chakraborti (1992) and Weitzman (1970) and references therein).

In this paper illustrative examples of relationships between stochastic dominance criteria and the notion of polarization are outlined and a taxonomy of states reflecting polarization both between and within distributions are described in Section 1. Section 2 presents some statistical tests for both between and within population polarization. Applications from the economic inequality, economic growth, union/gender wage effects and assortative mating literatures provide examples of tests for various types of polarization in Section 3. Section 4 draws some conclusions.

---

[2]See Johnson, Kotz and Balakrishnan (1994) on multimodality.

[3]For example the Wolfson (1994) argument that a population can simultaneously become more polarised and more equal under conventional measures of inequality is readily demonstrated by considering f(x) an equally weighted mixture of normals $N(\mu_i, \sigma_i^2)$, i = 1,2 and noting that $\sigma_x^2 = .5(\sigma_1^2+\sigma_2^2+(\mu_1-\mu_2)^2)$. Increased polarization, interpreted as any combination of reductions in sub population variances and increased divergence of their means, can be seen to either increase, leave unchanged or decrease inequality as measured by the variance of x.

1. Polarization and the relationship with stochastic dominance, some examples.

Working with a set of discrete income classes and a notion of proximity between those classes, Esteban and Ray (1994) focus upon an axiomatic description of within distribution polarization in order to develop a polarization index which reflects the incidence and intensity of multiple poles. Their first two axioms may be summarised as defining an increase in polarisation by partitioning the population into two groups, fixing one and requiring that the other reduce its dispersion without its location measure moving closer to the fixed group. The third axiom considers the reallocation from a central mass to two lateral masses with without causing their measures of central tendency to converge as increasing polarisation. To avoid difficulties associated with changes in population sizes a homotheticity assumption is added. Essentially to describe the phenomenon Esteban and Ray break down the distribution of interest into sub distributions and characterize Polarization in terms of the relative movements of these sub distributions[4].

Types of polarization reflecting these ideas in the context of continuous distributions entertaining the possibility of just two poles are illustrated in diagrams 1-4. Consider g(x) and f(x) as the respective distributions of characteristic x amongst two sub populations (say the poor club and the rich club) which are exclusive and exhaustive of the total population. When the club

---

[4] Their focus is the generation of a polarization index for the discretely distributed random variable y which takes on any one of n values y_i with probabilities π_i , i=1,..,n which, for constants K and α is of the form:

$$P(\pi,y) = K\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i^{1+\alpha}\pi_j|y_i-y_j|$$

K is simply a multiplicative constant which does not affect the ordering, α is a parameter reflecting the polarization sensitivity of the measure where $0< \alpha \le 1.6$, the larger its value the further is the measures departure from an inequality measure.

members are observed directly, between population comparisons can be made, when they are only observed as members of the overall population the observed distribution becomes a mixture $m(x) = \alpha g(x) + (1-\alpha) f(x)$ where α is the proportion of the total population in the poor club.

Diagrams 1- 3 consider limiting types of between population polarization and diagrams 1a-3a correspond to the within population polarization engendered by 50/50 mixtures of the separate sub populations represented in diagrams 1-3. Diagram 1 illustrates polarization (1st Order) created by allowing respective period 2 distributions $g^2(x)$ and $f^2(x)$ to correspond to lower and upper period 1 distributions $g^1(x)$ and $f^1(x)$ having suffered location shifts of -0.1 and 0.1 respectively, the poles have not increased in intensity but they have moved further apart and the extent to which the distributions overlap has diminished. Diagram 2 illustrates polarization (2nd Order) engendered by allowing $g^2(x)$ and $f^2(x)$ to be scale reduced versions of $g^1(x)$ and $f^1(x)$, the poles have increased in intensity but have not moved further apart however again the extent to which the distributions overlap has diminished. Yet a third polarization type (3rd Order) is demonstrated in diagram 3 by allowing $g^2(x)$ and $f^2(x)$ respectively to be mean and variance preserving left and right skewed transformations of $g^1(x)$ and $f^1(x)$. The lower g(x) distribution has moved toward its lower tail and the upper f(x) distribution has moved toward its upper tail[5].

The consequences for within distribution comparisons of 50/50 mixtures of $g^i(x)$ and $f^i(x)$ are illustrated in diagrams 1a - 3a where $m^i(x) = 0.5g^i(x) + 0.5f^i(x)$ for i = 1,2. In each case the mixture has become less dense at the center reflecting the reduction in the extent to which sub distributions

[5] For the between distribution comparisons the baseline period 1 distributions $g^1(x)$ and $f^1(x)$ are plots of Beta (3,3) distributed random variables with location shifts of +0.1 and +0.7 respectively. 2nd order Polarization was achieved by setting $g^2(x)$ and $f^2(x)$ to Beta(4,4) distributions with respective location shifts if +0.1 and +0.7. 3rd order Polarization was achieved by setting $g^2(x)$ to an appropriate location shifted linear combination of Beta(3,4) and Beta(2,3) variables and by setting $f^2(x)$ to an appropriate location shifted linear combination of Beta(4,3) and Beta(3,2) variables.

intersect on moving to period 2. In effect mass has been dispersed laterally from the center again in the spirit of Esteban and Ray, though note this does not mean the extreme tails necessarily extend (see for example diagram 2). In each case there is a sense in which the populations represented have become increasingly separated or polarized in the spirit of Esteban and Ray (1994), primarily characterized by a diminution in the extent to which respective "i" period distributions overlap in the case of between population comparisons and by central mass reduction in the case of mixtures.

It would seem that simple tests of central mass reduction in the case of mixtures would be appropriate for polarization detection, however this is not the case since difficulties are encountered when the initial sub-populations are very close together or very far apart. Diagrams 4 and 4a reflect the closeness problem in a variance reduction case comparable to diagrams 2 and 2a but with a sub population mean difference of only 0.05. Sub population polarization (variance reduction) now results in a relatively small diminution in the overlap measure and an increasing mass at the center of the mixture distribution[6] . Undoubtedly polarization has taken place, the sub-populations have separated further in a very real sense with the rich and poor clubs converging on their respective mean incomes, but when the clubs are not sufficiently far apart it manifests itself as a minimal reduction in overlap measure and an increase in mass at the center of the mixture with obvious consequences for polarization detection. When the sub population distributions are very far apart there will be little or no overlap to measure with any accuracy in the between population comparisons and little or no mass to change at the center of the corresponding mixture. In order to get round these problems tests need to be developed which reflect notions of changes in distance between distributions or the lower and upper portions of a mixture.

---

[6]Indeed it can be shown that, for this configuration of beta distributions in periods 1 and 2 and the variance reduction implied, the difference in means must be at least a 0.376 to engender a loss of mass at the center of the mixture. This phenomena is not unique to Beta distributions, for 50/50 mixtures of normals with a common σ and a proportionate variance reduction of $1-\alpha > 0$ the difference in means must be greater than $(2\sigma\alpha\ln\alpha/(\alpha^2-1))^{0.5}$ to effect a central loss of mass.

The idea that polarization has to do with population distributions (or sub population distributions in the case of mixtures) separating in some sense has a great deal in common with stochastic dominance criteria which are readily interpreted as describing the extent and type of distance between densities. Stochastic dominance conditions emerged in economic analysis in the context of portfolio choice (Hadar and Russell (1969)) and in the analysis of social welfare (Foster and Shorrocks (1986), Atkinson (1987)). In both cases they arise from contemplating the expected gain (either in terms of utility or social welfare) from moving from one distribution to another[7]. Consider δ, the difference in expected value of a function u(x) with the properties $(-1)^{j-1}\partial^j u/\partial x^j \geq 0$ j = 1,.,i for some i > 0 based upon two potential density functions F(x) and G(x) defined on the interval [a,b]. It may be written as:

$$\delta = E_F(u(x)) - E_G(u(x)) = \int_{-\infty}^{\infty} u(x)(dF - dG)$$

Necessary and sufficient conditions for δ > 0 for some given i are:

$$\int_a^x (F_{i-1}(z) - G_{i-1}(z))dz \leq 0 \textit{ for all } x \quad [1]$$

and

$$\int_a^x (F_{i-1}(z) - G_{i-1}(z))dz < 0 \textit{ for some } x \quad [1a]$$

where, letting $f(x) = F_0(x)$, $F_i(x)$ is defined recursively as:

---

[7]The following arguments apply equally to discrete and continuous random variables, for brevity they are explored in the continuous paradigm though a discrete example will be presented.

$$F_i(x) = \int_a^x F_{i-1}(z)dz: \quad (a \le x \le b,\ i \ge 1)$$

with $G_i(x)$ defined similarly. When [1] and [1a] are satisfied f (x) stochastically dominates g(x) at order i. In the following $f(x) \succeq_j g(x)$ denotes compliance with [1] at least order j, $f(x) \succ_j g(x)$ denotes compliance with [1a] at order j and $f(x) \succ\succ_j g(x)$ denotes that the strict inequality in [1a] obtains over a relevant range. Note that for $i < j$, $f(x) \succeq_i g(x)$ implies $f(x) \succeq_j g(x)$, furthermore it is a transitive relationship in that if $f(x) \succeq_j g(x)$ and $g(x) \succeq_j h(x)$ then $f(x) \succeq_j h(x)$.

A convenient interpretation of these conditions is in terms of the degree of "Right Separation" between the two distributions. When $f(x) \succeq_i g(x)$, $F_i(x_1) = G_i(x_2)$ implies $x_2 \le x_1$, so that G_i is everywhere **not** to the right of F_i and, if [1a] holds at level i, to the left of it at least somewhere, implying Right Separation of f(x) from g(x) at the i'th level of integration. As limiting examples let x be a transformation of y with resultant respective pdf's $f^2(\ )$ and $f^1(\ )$, then:

1) a positive location shift transformation implies $f^2(x) \succeq_1 f^1(x)$ and $f^2(x) >>_1 f^1(x)$
2) a location preserving, scale reducing transformation implies $f^2(x) \succeq_2 f^1(x)$ and $f^2(x) >>_2 f^1(x)$
3) a location and scale preserving, positive skewing shift implies $f^2(x) \succeq_3 f^1(x)$ and $f^2(x) >>_3 f^1(x)$

which correspond to the $f^1(\ )$ to $f^2(\ )$ distribution transformations in Diagrams 1 - 3 respectively.

Of equal of interest is a similar notion of "Left Separation" where a conditions of the form:

$$\int_x^b (F^{i-1}(z) - G^{i-1}(z))dz \le 0 \text{ for all } x \quad [2]$$

and

$$\int_x^{\infty} (F^{i-1}(z) - G^{i-1}(z))dz < 0 \text{ for some } x \quad [2a]$$

where, letting $f(x) = F^0(x)$, $F^i(x)$ is defined recursively as:

$$F^i(x) = \int_x^b F^{i-1}(z)dz: \quad (a \leq x \leq b,\ i \geq 1)$$

and $G^i(x)$ is defined similarly. This type of dominance is employed in the analysis of risk-loving behaviour characterized by preferences with properties $\partial^j u/\partial x^j \geq 0$ j = 1,.,i for some i > 0 (see Levy and Weiner (1998)). Consider a transformation of x of the form w = -x where f(w) and g(w) are the suitably transformed distribution functions now defined over [-b,-a] then conditions [2] and [2a] are equivalent to $f(w) \succeq_i g(w)$ and $f(w) \succ_i g(w)$ respectively and have the analogous "Left Separation of f(x) from g(x)" interpretation. The progress from $g^1(x)$ to $g^2(x)$ in diagrams 1 - 3 corresponds to the first three degrees of left separation.

Thus overall Diagram 1 corresponds to $f^1(x) \succeq_1 g^1(x)$, $g^2(w) \succeq_1 g^1(w)$ and $f^2(x) \succeq_1 f^1(x)$ whereas in Diagrams 2 and 4 $f^1(x) \succeq_1 g^1(x)$, $g^2(w) \succeq_2 g^1(w)$ and $f^2(x) \succeq_2 f^1(x)$ whilst in Diagram 3 $f^1(x) \succeq_1 g^1(x)$, $g^2(w) \succeq_3 g^1(w)$ and $f^2(x) \succeq_3 f^1(x)$. In each case the dominated g( ) distribution has left separated and the dominant f( ) distribution has right separated in some sense. Clearly tests which establish the prevalence of these criteria would be suitable candidates for polarization tests and preferable to simple overlap comparisons.

Two further ideas will be useful in employing dominance criteria to study polarization. The first involves considering dominance over a subset of [a,b] in an obvious fashion. An important theorem, useful in what follows, is provided as Lemma 1 in Davidson and Duclos (2000) and will subsequently be referred to as the Davidson-Duclos theorem. In essence if f(x) $\succ\succ_1$ g(x) for x $\in$ [a,z] with a < z < b, then f(x) $\succeq_s$ g(x) for all x $\in$ [a,b] for s sufficiently large. Thus if first order dominance can be established at the low end of a distribution, there exists some order of dominance s which pertains over the complete domain of the distribution. Alternatively the existence of some order of dominance s over the complete domain implies strict first order dominance for some region [a, z].

The second idea involves the dominance of differences. Consider four distributions $f^1(x)$, $f^2(x)$, $g^1(x)$ and $g^2(x)$ and contemplate δ_D, the excess in the expected gain when moving from $f^1(x)$ to $f^2(x)$ over that of moving from $g^1(x)$ to $g^2(x)$. It may be written as:

$$\delta_D = [E_{f^2}(u(x)) - E_{f^1}(u(x))] - [E_{g^2}(u(x)) - E_{g^1}(u(x))]$$
$$= \int_a^b u(x)((f^2(x) - f^1(x)) - (g^2(x) - g^1(x)))dx$$

Letting $f^d(x) = f^2(x) - f^1(x)$ and $g^d(x) = g^2(x) - g^1(x)$ necessary and sufficient conditions for $\delta_D > 0$ are simply the compliance of $f^d(x)$ and $g^d(x)$ with [1] and [1a] respectively which for notational convenience will be denoted as f(x) D_i g(x) and referred to as "i'th Order Dominance of Distributional Differences".

Interest in this concept can be motivated by contemplating the extent to which the overlap of two population distributions is reduced[8]. Consider two populations with smooth continuous unimodal overlapping pdf's g(x) and f(x) defined on the real line such that f(x) $\succeq_1$ g(x) and define

---

[8] Weitzman (1970) employs an overlap measure in an analysis of discrimination.

$x^* = x \mid g(x)=f(x) > 0$ and note that $f(x) < g(x) \mid x < x^*$ and $f(x) > g(x) \mid x > x^*$. The extent to which the distributions overlap is given by:

$$OV = \int_{-\infty}^{\infty} \min(f(x), g(x))dx = \int_{-\infty}^{x^*} f(x)dx + \int_{x^*}^{\infty} g(x)dx$$

Letting superscript i denote state i , the reduction in OV between states 1 and 2 may be written:

$$OV^1 - OV^2 =$$

$$\int_{-\infty}^{x^{1*}} (f^1(x) - f^2(x))dx + \int_{x^{2*}}^{\infty} (g^1(x) - g^2(x))dx + \int_{x^{1*}}^{x^{2*}} (g^1(x) - f^2(x))dx \qquad for\ x^{1*} \leq x^{2*} \qquad [1a]$$

$$\int_{-\infty}^{x^{2*}} (f^1(x) - f^2(x))dx + \int_{x^{1*}}^{\infty} (g^1(x) - g^2(x))dx + \int_{x^{2*}}^{x^{1*}} (g^1(x) - f^2(x))dx \qquad for\ x^{1*} \geq x^{2*} \qquad [1b]$$

Proposition: $(f^1(x)-f^2(x)) \leq (g^1(x)-g^2(x))$ for all $x < x^{1*}$ with strict inequality holding for some x in that region $\Rightarrow OV^1 - OV^2 > 0$. (For proof see Appendix)

Proof. Since:

$$\int_{x^*}^{\infty} g(x)dx = 1 - \int_{-\infty}^{x^*} g(x)dx$$

write [1a] as:

$$OV^1 - OV^2 = \int_{-\infty}^{x^{1*}} (f^1(x) - f^2(x))dx + \int_{-\infty}^{x^{2*}} (g^2(x) - g^1(x))dx + \int_{x^{1*}}^{x^{2*}} (g^1(x) - f^2(x))dx$$

$$= \int_{-\infty}^{x^{1*}} ((f^1(x) - f^2(x)) - (g^1(x) - g^2(x)))dx + \int_{x^{1*}}^{x^{2*}} (g^2(x) - g^1(x))dx + \int_{x^{1*}}^{x^{2*}} (g^1(x) - f^2(x))dx$$

$$= \int_{-\infty}^{x^{1*}} ((f^1(x) - f^2(x)) - (g^1(x) - g^2(x)))dx + \int_{x^{1*}}^{x^{2*}} (g^2(x) - f^2(x))dx \qquad [2a]$$

When $x^{1*} \leq x^{2*}$, $g^2(x) \geq f^2(x)$ for all $x \in [x^{1*}, x^{2*}]$ and the last integral in [2a] is non-negative. By similar manipulation [1b] may be written as:

$$OV^1 - OV^2 = \int_{-\infty}^{x^{1*}} ((f^1(x)-f^2(x))-(g^1(x)-g^2(x)))dx + \int_{x^{2*}}^{x^{1*}} (f^2(x)-g^2(x))dx \qquad [2b]$$

When $x^{1*} \geq x^{2*}$, $f^2(x) \geq g^2(x)$ for all $x \in [x^{2*}, x^{1*}]$ and the last integral in [2b] is non-negative. The condition of the proposition guarantees that the first integral common to both [2a] and [2b] is strictly positive in each case hence $OV^1 > OV^2$.□

Intuition is gained by rewriting the strictly positive common first term in [2a] and [2b] as:

$$\int_{-\infty}^{x^{1*}} (f^1(x)-f^2(x))dx + \int_{x^{1*}}^{\infty} (g^1(x)-g^2(x))dx > 0 \qquad [3]$$

As long as the reduction in the area of overlap below x^{1*} exceeds the increase in it above x^{1*}, a reduction in overlap will occur. First order dominance of differences up to x^{1*} guarantees this, it is not necessary but it is a sufficient condition. A stronger version of this condition would be where f^2 first order dominates f^1 and g^1 first order dominates g^2 up to x^{1*} making both terms in [3] positive securing overlap reductions both below and above x^{1*}. This is much more than is necessary to establish an overall reduction in overlap and can be thought of as an extreme or strong form of Polarization as opposed to the weaker form of dominance of differences. This leads to four possible forms of between population polarization.

A Taxonomy Polarization Between two distinct populations.

Changes in relative population sizes obfuscate distributional aspects of polarization (hence ER's homotheticity lemma) thus population sizes are assumed constant in what follows.

Fundamentally polarization with respect to a characteristic in two populations has to do with their respective distributions moving further apart in some sense. Let $f^j(x)$ and $g^j(x)$ represent distributions f and g in period j corresponding respectively to two populations.

Strong i'th Order Polarization.

Strong polarization requires that two sets of conditions simultaneously pertain:

Sa) $f^1(x) \succeq_1 g^1(x)$ and $f^1(x) \succ_1 g^1(x)$.

Sb) $f^2(x) \succeq_i f^1(x)$, $g^2(w) \succeq_i g^1(w)$ and either $f^1(x) \succ_i f^2(x)$ and / or $g^1(w) \succ_i g^2(w)$

Sa) ensures that initially F(x) is everywhere to the right of G(x) establishing the notion that f(x) and g(x) are apart and the direction of the separateness. Given the transitivity property of dominance relations, Sb) ensures that the inter-temporal changes result in f and g now being further apart or more polarised. This is so since neither distribution has moved closer to the other and either the dominant distribution has Right Separated and / or the dominated distribution has Left Separated in some sense.

Semi-Strong Polarization.

SSa) $f^1(x) \succeq_1 g^1(x)$ and $f^1(x) \succ_1 g^1(x)$.

SSb) $f^2(x) \succeq_1 f^1(x)$, $g^2(w) \succeq_1 g^1(w)$, $f^1(x) \succ_1 f^2(x)$ and $g^1(w) \succ_1 g^2(w)$ for $x \in [a,z]$, $z<b$ and $w \in [-b,y]$, $y<-a$.

Again SSa) establishes that f(x) and g(x) are apart together with the direction of their separateness. Given the transitivity property of dominance relations and using the Davidson-Duclos theorem, SSb) establishes that f and g are now further apart or more polarised since f has Right Separated and g has Left Separated, however now the order of dominance under which they have moved is of an unknown order s.

These may be overly stringent criteria since they rule out situations where $g^2(x)$ Right Separates from $g^1(x)$ but the extent to which $f^2(y)$ Right Separates from $f^1(y)$ more than compensates for this. For this situation resort is made to the definition of Weak Polarization.

Weak i'th Order Polarization.

Two conditions are required for weak polarization to obtain:

Wa) $f^1(x) \succeq_1 g^1(x)$ and $f^1(x) \succ_1 g^1(x)$.

Wb) $f(x)\ D_i\ g(x)$

Again Wa) plays its usual role whilst Wb) establishes that f and g are now further apart or more polarised since f has moved more further to the right than g has in some sense. Obviously Strong Polarization of a given order implies Weak Polarization of the same order.

Super Weak Polarization.

In a natural extension to weak polarization that provides an analogue to semi strong polarization two conditions are required to hold simultaneously.

SWa) $f^1(x) \succeq_1 g^1(x)$ and $f^1(x) \succ_1 g^1(x)$.

SWb) $f(x)\ D_1\ g(x)$ for $x \in [a,z]$, $z<b$.

Polarization Within a Population.

Polarization within a population distribution concerns a tendency toward bi-modality characterised by a loss of mass at the centre of the distribution. Unfortunately such tendencies may exist long before multiple modes emerge[9] , thus modification of tests for the existence of bi-modality are unlikely to prove fruitful. The possibility that changes in the underlying structure of distributions in the mixture are the source of the phenomenon has to be explored. Indeed applying stochastic dominance criteria to mixture distributions whose mixing parameters are constant results in the same mixtures of the corresponding dominance criteria. Consider simple mixtures of the form $f(x) = \theta f^r(x) + (1-\theta)\, f^p(x)$ and $g(x) = \zeta g^r(x) + (1-\zeta)g^p(x)$ and let $\theta = \zeta + \gamma$, it is readily shown that:

$$F_i(x) - G_i(x) = \theta(F^r_i(x)-G^r_i(x)) + (1-\theta)(F^p_i(x)-G^p_i(x)) + \gamma(G^r_i(x)-G^p_i(x))$$

so that when the mixing proportions remain constant ($\gamma = 0$) dominance criteria for mixtures are the corresponding mixtures of the dominance criteria.

Identification of the sub-population empirical distribution functions is not possible directly, however some qualified inferences can be made by associating the lower and upper parts of the observed mixture with the respective unobserved dominated and dominating distributions in the mixture. However caution is to be advised when recalling from section 1 the problems engendered by sub populations being located too close together initially. Partitioning the distributions at some common defining point[10] x^* and considering the relative progress of the distributions $f^1(x|x<x^*)$, $f^2(x|x<x^*)$, $f^1(x|x>x^*)$ and $f^2(x|x>x^*)$ along the lines suggested above for the between population comparisons can be illuminating. Clearly $f^1(x|x<x^*) \leq_1 f^1(x|x>x^*)$ is always true in this case and

[9] For example, for mixtures of normals with equal variances, bi-modality will not emerge under any mixing scheme until the difference in means exceeds $4.5^{0.5}$ standard deviations (Johnson, Kotz and Balakrishnan (1994) p 164).

[10] The context is important here, generally it may be the mean or the median, in income distribution analysis it may be the poverty level, in educational attainment studies it may be a pre-assigned level of schooling.

does not need to be established so that the "a)" condition employed in the two population cases to establish the direction of separateness of the initial distributions is no longer required.

What do the different orders of inter-temporal dominance imply for polarity? $\succeq_3$ requires at least mean difference and variance preserving transfers from the centre toward the poles (Diagrams 3 and 3a), $\succeq_2$ requires at least mean difference preserving distribution shrinkage at the poles (Diagrams 2 and 2a) and $\succeq_1$ requires that the poles move apart without any increase in scale around the poles (Diagrams 1 and 1a).

2. Tests for Polarization.

Different forms of polarization can be examined empirically via jointly testing various collections of dominance relationships. Tests for Stochastic dominance relationships have proliferated in the literature in recent years by employing the distribution of integral approximations (Anderson (1996)), by employing the distribution of incomplete moments (Davidson and Duclos (2000)) or by employing the distribution of functions of the empirical distribution function (McFadden (1989), Barrett and Donald (1999)). The last family of tests are attractive because, unlike the first two families, they are consistent tests, however it can be shown that, under smoothness assumptions, the inconsistency problem is not substantive (Anderson (2001b)). The first two families of tests are attractive because they are easily adapted to situations where samples are non i.i.d. (Anderson (1998), Davidson and Duclos (2000)). Here the integral approximation approach is employed.

All of the above techniques involve the joint testing of inequality restrictions. The tests for stochastic dominance based upon independent sampling derived in Anderson (1996) are based upon linear transformations of $p^f - p^g$, each k long vectors of observation proportions falling into k mutually exclusive and exhaustive predetermined intervals defined on the domain of the two distribution functions. Letting d_j be the j'th interval length then, were F known, probabilities of falling in the j'th category would be given by:

$$p_j = F(y^j) - F(y^{j-1}); \; where \; y^h = \sum_{i=1}^{h} d_i; \; F(y^0) = 0;$$

Adjusting the trapezoidal rule for approximating integrals (Goodman (1967)) to permit non equal intervals suggests that:

$$F(y^j) = \sum_{i=1}^{j} p_i \; ;$$

$$C(y^j) = \int_0^{y^j} F(z)dz \simeq .5[F(y^j)d_j + \sum_{i=1}^{j-1} (d_i + d_{i+1})F(y^i)];$$

$$\int_0^{y^j} C(z)dz \simeq .5[C(y^j)d_j + \sum_{i=1}^{j-1} (d_i + d_{i+1})C(y^i)].$$

by defining two matrices as follows:

$$I_f = \begin{Vmatrix} 1 & 0 & 0 & . & . & 0 \\ 1 & 1 & 0 & . & . & 0 \\ 1 & 1 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ 1 & 1 & 1 & . & . & 1 \end{Vmatrix}$$

$$I_F = .5 \begin{Vmatrix} d_1 & 0 & 0 & . & . & 0 \\ d_1+d_2 & d_2 & 0 & . & . & 0 \\ d_1+d_2 & d_2+d_3 & d_3 & . & . & 0 \\ . & . & . & . & . & . \\ d_1+d_2 & d_2+d_3 & d_3+d_4 & . & . & d_k \end{Vmatrix}$$

discrete analogues allow contemplation of i'th order dominance by focussing on:

$$H_0 : (I_F)^{i-1} I_f (p^A - p^B) \geq 0 \textit{ against } H_1 : (I_F)^{i-1} I_f (p^A - p^B) < 0:$$

$$\textit{note } H_2 : (I_F)^{i-1} I_f (p^A - p^B) \nleq \wedge \ngeq 0 \Rightarrow \textit{indeterminate.}$$

where in each case under the alternative strict inequality must hold for at least one element of the vector. Largely speaking the tests can be formed by stacking the appropriate comparison vectors and establishing their distributions. Letting p^{hj} be the probability vector for population h (= A,B) in period j (=1,2).Having partitioned the range of a random variable x into k mutually exclusive and exhaustive categories then the number of observations on x falling in the i'th category, denoted y_i, is asymptotically distributed multinomial with probabilities π_i, i = 1,..,k. It follows that $\sqrt{n}(p^{hj} - \pi^{hj}) \sim N(\pi^{hj}, \Omega^{hj})$ where $\Omega^{hj} = \mathrm{diag}(\pi^{hj}) - \pi^{hj} (\pi^{hj})'$. Assuming sampling to be independent across populations and time periods and for notational convenience assume the sample sizes to be identical (adjustments for different sample sizes are simple to implement see Anderson (1996)) then $\sqrt{n}(p^{hj} - p^{qr}) \sim N(\pi^{hj} - \pi^{qr}, \Omega^{hj} + \Omega^{qr})$ and the various dominance hypotheses can be examined via $v_i(h,j,q,r) = \sqrt{n} I_F^{i-1} I_f (p^{hj} - p^{qr}) \sim N(I_F^{i-1} I_f (\pi^{hj} - \pi^{qr}), (I_F^{i-1} I_f)(\Omega^{hj} + \Omega^{qr})(I_F^{i-1} I_f)')$.

Polarization hypotheses can be examined by stacking the appropriate vectors $v_i(h,j,q,r)$ taking care to compile the appropriate covariance matrix. The only abnormal covariance component in this context is that of p corresponding to g(x) and p corresponding to g(-x). Lett the former be Ω with typical term ω_{ij} and the latter be Ω_{rev}, the two matrices can be shown to obey the relationship:

$$\Omega = \begin{Vmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \cdot & \cdot & \omega_{1k} \\ \omega_{21} & \omega_{22} & \omega_{23} & \cdot & \cdot & \omega_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \omega_{k1} & \omega_{k2} & \omega_{k3} & \cdot & \cdot & \omega_{kk} \end{Vmatrix} ; \ \Omega_{rev} = \begin{Vmatrix} \omega_{1k} & \omega_{1k-1} & \omega_{1k-2} & \cdot & \cdot & \omega_{11} \\ \omega_{2k} & \omega_{2k-1} & \omega_{2k-2} & \cdot & \cdot & \omega_{21} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \omega_{kk} & \omega_{kk-1} & \omega_{kk-2} & \cdot & \cdot & \omega_{k1} \end{Vmatrix}$$

The practice has been to employ either the Maximum Modulus Distribution (tables for which are provided in Stoline and Ury (1978)) to a collection of asymptotically normal statistics (which is a conservative test) or to employ the joint testing procedures advocated in Kodde and Palm (1986) and Wolak (1989). The advantage of the former is that the inequality relations may be studied in detail, the advantage of the latter, it is not a conservative test. The conservative tests can be pursued by considering the appropriately stacked vectors of v_i and the corresponding elements of their covariance matrix. For the joint test let v_{iw} be the inequality constrained estimate of the vector v_i and let $(\Omega_{vi})^+$ be the Moore- Penrose inverse of the covariance matrix of v_i, then for:

$$W = (v_i - v_{iw})'(\Omega_{v_i})^+(v_i - v_{iw})$$

the distribution of W is such that:

$$P(W \geq c) = \sum_{i=0}^{k-1} P(\chi^2_i \geq c) w(k,k-i,\Omega_{v_i})$$

where $w(k,k-i,\Omega)$ is a weight function corresponding to the probability that v, with covariance matrix Ω, has k-i-1 of its k-1 independent elements positive. This weight function is complex to compute and closed form expressions only exist for k-1 up to 4, however following the suggestion in Wolak (1989) they can readily be approximated via pseudo normal random number generation.

3. Empirical Examples.

Four empirical examples illustrate the foregoing ideas. The first, from inequality and poverty analysis, exemplifies the application of the strong polarization criteria between continuous populations. The second, from the economic growth literature, provides an example of within population weak polarization criteria and the third from a study of the impact of trade unions on wage distributions exemplifies the between population weak polarization criteria both again for continuous populations. The fourth, drawn from the Assortative Mating by Educational Attainment literature, illustrates the application of within population strong polarization criteria in the context of discrete distributions. In the following tables the upper tail probabilities of the null hypothesis employing Wolak's multiple comparison procedures are presented for the first three examples, the last example illustrates the use of the more conservative - but perhaps more informative - maximum modulus distribution type tests.

3.1. The plight of single parent mothers.

Throughout the 1980's in Canada the plight of single parent mothers was of considerable concern since it is considered the primary source of child poverty. The sense is that, compared to other family units, female single parent headed units relative economic status was deteriorating, exacerbating the plight of children in those units. The question of whether they were being economically polarized can be addressed in the following context. The progress of Single parent mothers (SPM's) with one child are compared to single females (SF's) in terms of their real after tax incomes (deflated in the former case by $\sqrt{2}$ to account for an adult equivalent income). Between the years 1981 and 1989. 1981 and 1989 mean incomes for SPM's were C$11849.1 and C$11531.3 respectively, the corresponding magnitudes for SF's were C$13258.6 and C$13779.9. It is evident from Table 1 that Single Females are initially better off than Single Mothers and that whilst the position of SF's improved (the '89 distribution 2nd order dominated the '81 distribution) the position of the SPM's deteriorated (the '81 distribution 1st order dominates the '89 distribution at the 1% level and 2nd order dominates it at any other usual significance level. It is then no surprise from Table 2 that the Strong Form 2nd Order Polarization is indicated for SPM's and that the relative plight of those units did indeed deteriorate.

Table 1. Stochastic Dominance Comparisons.

Comparison	Test	A dominates B	B dominates A
Single females (A) versus single parent mothers (B) 1981	1st order	0.524	0
	2nd order	0.621	0
	3rd order	0.583	0
Single females 1981(A) versus 1989(B)	1st order	0	0
	2nd order	0	0.665
	3rd order	0	0.624
Single parent mothers 1981(A) versus 1989(B)	1st order	0.045	0.008
	2nd order	0.683	0
	3rd order	0.641	0

Table 2. Polarization Tests.

Test	Order	Polarization	Non-Polarization
Strong form	1st	0	0
	2nd	0.382	0
	3rd	0.379	0
Semi-Strong form		0.741	0

3.2. The Convergence Hypothesis in Economic Growth.

It has been well established in the endogenous growth literature that GDP growth rates for different societies conditionally converge. Usually the convergence rate is assumed constant across countries, though there is evidence from between educational attainment and per capita GDP interaction terms that less educated societies have slower convergence rates (see for example Barro(1998), Jones (1997) and Quah (1997)) and hence different conditional distributions of growth rates. Diffusion theories predicated upon imitation being cheaper than innovation also suggest that these convergence rates should themselves converge in the long run. It is important for poverty comparisons of poor versus rich nations to establish whether this maps into unconditional convergence the GNP growth rates. The question is that, whilst less developed and developed countries form two distinct groups characterised by their growth rate distributions, are they converging to one group over time through the diffusion process? This has a natural interpretation as the reverse of polarization or depolarization and should manifest itself as such in the comparison of the distribution across nations of growth rates over long intervals of time.

The distribution is of necessity a mixture of less developed and developed nation growth rates and, given the observation in the introduction, convergence is unlikely to be captured by a simple variance reduction test. Within population weak polarization is examined with, for comparison, the "F" and two non-parametric variance ratio tests (Mood(1954) and Klotz(1962) which can be shown to have preferable properties under non normality Anderson (2001a)). Data for per capita growth rates of GNP for various economies is derived from the World Bank Development Indicator Tables 1997 for 6 comparisons 20 years apart starting in 1970. Here the null of weak convergence is the null of polarization in the initial year relative to its comparator so that, for example, under the null 1970 would be polarized relative to 1990.

Observe from Table 3 that, in spite of the relatively small samples involved, weak convergence is supported at the 1% critical value for 5 of 6 comparisons and it is supported at any other usual critical value for three of the comparisons. Divergence is never supported in any comparison, however the caveat issued in Section 1 regarding sub population means being too close together should be recalled and unqualified conclusions resisted accordingly. It is of interest to note that the various variance ratio tests support the convergence hypothesis in only one case (1975-1995) reflecting the difficulty in identifying or associating polarization with an increased variance.

Table 3 Convergence Tests.

Years	Weak Convergence		Super-Weak Convergence		Scale Tests (for null of no variance diminution)			Sample Sizes
	H_0	H_1	H_0	H_1	"F"	Mood	Klotz	
1970/90	0	0	0.068	0	0.351	0.747	0.324	114/163
1971/91	0.012	0	0	0	0.997	0.811	0.907	122/162
1972/92	0.232	0	0.336	0	0.999	0.284	0.3	122/161
1973/93	0.981	0	0.809	0	0.757	0.131	0.473	122/160
1974/94	0.011	0	0	0	0.828	0.22	0.262	126/157
1975/95	0.928	0	0.828	0	0	0.01	0.01	127/155

3.3. Female versus Male unionization wage distribution effects.

That unions increase the location and compress the scale of wage distributions has been well documented (Dinardo and Lemieux (1997)) and that in Canada the process has been more exaggerated for females than for males (Doiron and Riddell (1994)) leads to many possibilities for examining polarization effects in wage distributions. To examine these possibilities data on wage rates for the years 1993 and 1994 were culled from the Survey of Labour and Income Dynamics. Separate comparisons were made between unionized and non-unionized full-time workers for male and female categories, details of the raw data are presented in Table 4. Given the upward drift of wages for all categories of workers, it is not surprising that strong polarization was rejected for all comparisons. Table 5 presents the dominance and weak polarization test results for male and female workers.

Table 4. Union/Non Union Wage Rates

	1994			1993		
	Mean	St. Dev.	N	Mean	St. Dev.	N
Male Union	19.767	6.238	2496	19.524	6.284	2350
Male Non-Union	20.293	8.702	1509	20.246	8.534	1448
Female Union	17.667	6.325	1919	17.394	6.168	1789
Female Non-Union	15.095	6.898	1079	14.723	6.737	1063

With regard to male workers they indicate that, although the year by year changes suggest an improvement in the union wage distribution over the non-union distribution, since 1st order union wage distribution dominance could not be established in the initial year, no polarization effects could be established for males. A different story emerges for female workers. With first order dominance of both the initial position and year by year changes favouring the union over non-union wage distributions, first order weak polarization of female union vs non-union wage distributions is readily established. It is surprising that such effects, though expected over the longer run, can be so clearly observed in contiguous observation years.

Table 5. Dominance and Polarization Tests (Males and Females).

Test	Order	A v B	B v A
Union (A) v. Non-Union (B) Males '93 Dominance test	1st	0	0
	2nd	0	0
	3rd	0.596	0
'93-'94 changes Union (A) v. Non-Union (B) Males Dominance test	1st	0.814	0
	2nd	0.65	0
	3rd	0.589	0
Males Weak Polarization test	1st	0	0
	2nd	0	0
	3rd	0	0
Super Weak	1st	0	0
Union (A) v. Non-Union (B) '93 Females Dominance test	1st	0.822	0
	2nd	0.655	0
	3rd	0.594	0
'93-'94 changes Union (A) v. Non-Union (B) Females Dominance test	1st	0.823	0
	2nd	0.647	0
	3rd	0.592	0
Females Weak Polarization test	1st	0.826	0
	2nd	0.785	0
	3rd	0.775	0
Super Weak	1st	0.724	0

3.4 Assortative Mating.

The correlation between spouses educational levels has attracted the interest of Sociologists (Mare (1991) and references cited therein), Demographers (Qian (1998) and references cited therein) and Economists (Becker (1981) Pencavel (1998), Mancuso and Pencavel (1999)) alike. Qian cites the Becker complementarity view of spousal roles as arguing for a negative tendency in the correlation between educational attainment levels. Consideration of the spousal contribution of the present value of future income streams to the partnership suggests a preference for mates with the highest possible educational attainment level and thus a positive correlation. In the context of a society where distributions of educational attainment by gender have both shifted higher and become more similar, discrimination between the theories becomes difficult since spousal educational attainment levels will become increasingly correlated in all cases. Assortative mating by educational status has been modeled in the style of the Gale and Shapley (1962) assignment game (see for example Mancuso and Pencavel (1999)). In its simplest form, the marriage market contains equal numbers of each gender, every gender member has the same strict ordering of preferences over all members of the opposite gender, including the notion that marriage of any sort is preferable to being single.

When agents prefer mates with higher educational attainment per se, the core of the assignment game contains couples matched by rank (see Roth and Sotomayor (1990)), or positive assortative mating. Beckers model cannot be characterized in the same way since preference orderings over the opposite gender are not identical for all members of a given gender (an individuals preferences will depend upon their educational attainment level). However an extreme form of the Becker model (where greater spousal educational differences imply greater returns to complementarity engendered by specialisation) would, under the assumption of identical educational attainment medians for both genders, permit appropriately ordered preferences of all above median males (females) over all below median females (males) resulting in a core containing paired reversed ranks or negative assortative mating. Under positive assortative mating the distribution of the difference in educational attainment between spouses would thus be dense at the median and thin at the tails whereas under negative assortative mating it would be dense at the tails and thin at the median. Thus a tendency toward negative assortative mating would be reflected in within population polarization of the differences in educational attainment with a tendency in the opposite direction implying depolarization.

Following Pencavel (1998) data for married or cohabiting couples are taken from the 1940, 1960 and 1990 Censuses of Population and schooling attainment is ranked into five educational categories, 1 - No High School, 2 - No college 3 - No More than 1 year of College, 4 - No More than 4 years of College and 5 - More than 4 years of college[11]. Table 6 presents the raw details by year and gender implied by these rankings. As may be seen females had higher average educational attainment than males in the 1940 sample (largely because there were far fewer of them in the "No High School" category), though this was reversed in subsequent sample years. Worthy of note is the general growth in mean educational attainment evident in both genders and the persistently lower variance in female attainment levels relative to males. Most importantly the positive correlation of spouses attainment levels implied by the juxtaposing the variances of male, female and spousal differences (indication positive assortative mating) diminishes over the period.

The Polarization Tests of Spousal Educational Attainment Differences are reported in Table 7 (the categories at the extremes were augmented because of the sparseness of the -4 and 4 categories), the 1940-1960, 1960-1990 and 1940-1990 tests are reported as the cell differences

[11]The manner in which educational attainment was recorded changed substantially between the 1960 and 1990 samples, however this breakdown affords a reasonably consistent categorization over the three observation periods.

and their asymptotic standard errors, the ratios of which are distributed as the Studentized Maximum Modulus distribution. Polarization is established by the presence of at least one significantly non positive cell and no significantly positive cells in a comparison. First order comparisons are limited to 3 cells, the fourth being redundant because of the adding up property of cumulative distributions. Upper .05 points for the above comparisons are 2.388 for three cell comparisons and 2.631 for four cell comparisons. Table 7 indicates that first order polarization occurred from 1940 to 1960 and from 1940 to 1990. For the 1960-1990 comparison a significantly positive third cell difference indicates a rejection of First Order Polarization, though the significantly negative values of the first two cells indicate by the Davidson-Duclos theorem dominance at some order. In fact Second Order Polarization was also rejected (the results are not reported here but are available from the author) however as the last two lines of the table indicate Third Order Polarization can be established between 1960 and 1990. Taken together these results suggest that whilst Assortative mating is positive, its extent has been weakening systematically over the last half of the century, contrary to assertions in the literature (see for example Mare (1991) with a tendency toward polarization.

Table 6 Sample Details.

	1940	1960	1990
Male average attainment	2.04008	2.62206	2.76139
Male attainment standard deviation	1.04883	1.19961	1.15477
Female average attainment	2.06763	2.60196	2.67122
Female attainment standard deviation	0.95697	1.02735	1.03604
Average difference in attainment	-0.02754	0.02010	0.09018
Difference in attainment standard deviation	0.86930	1.00778	1.00160
Implied attainment co-variance	0.63008	0.73945	0.70184
Implied attainment correlation	0.62776	0.59999	0.58663

Table 7 Polarization Tests.

Comparison		cell 1	cell 2	cell 3	cell 4	1st and last year Esteban - Ray Indices for α = 0, 0.5, 1.0 and 1.5			
1940-1960	v_i	-0.1499	-0.0863	-0.0882		0.8310	0.3984	0.2408	0.1644
(1st Order)	s.e.	0.0040	0.0034	0.0040		1.0421	0.4752	0.2550	0.1508
1960-1990	v_i	-0.0294	-0.0159	0.0755		1.0421	0.4752	0.2550	0.1508
(1st Order)	s.e.	0.0071	0.0066	0.0072		1.0297	0.4687	0.2552	0.1551
1940-1990	v_i	-0.1793	-0.1022	-0.0127		0.8310	0.3984	0.2408	0.1644
(1st Order)	s.e.	0.0070	0.0065	0.0073		1.0297	0.4687	0.2552	0.1551
1960-1990	v_i	-0.0073	-0.0334	-0.0558	-0.0444	1.0421	0.4752	0.2550	0.1508
(3rd Order)	s.e.	0.0018	0.0078	0.0181	0.0312	1.0297	0.4687	0.2552	0.1551

Conclusions.

There are many circumstances in economics (and no doubt in other disciplines) where issues have a polarization - depolarization interpretation both within and between populations. It is also evident that the essence of these phenomena cannot be captured by looking at standard measures of inequality movements such as variance related changes. To facilitate analysis a taxonomy of polarization types for between and within population comparisons has been provided in the spirit of Estaban and Ray (1994) together with a means of testing for their existence. Polarization is seen to have a natural and simple characterization and interpretation in the context of stochastic dominance relations between distributions or regions of those distributions under the assumption of constant population size[12]. Problems exist for the identification of polarization within mixture distributions when the means of the sub populations in the mixtures are not sufficiently far apart. Four examples demonstrate the pertinence of characterising issues in terms of polarization, the effectiveness of the comparison procedures in detecting the nature of polarization in a wide range of contexts and, given techniques for investigating stochastic dominance relations, the simplicity with which the comparisons can be executed.

References.

Anderson G.J., (1996) "Nonparametric Tests for Stochastic Dominance in Income Distributions" Econometrica 64 pp.1183-1193.

Anderson G.J. (1998) Poverty in America 1970-1990: Who Did Gain Ground? Mimeo University of Toronto Economics Department.

Anderson G.J. (2001a) "The power and size of Nonparametric Tests for Common Distributional Characteristics" Econometric Reviews 20, 1 - 30.

Anderson G.J. (2001b) "On the inconsistency of Tests Employing Point-wise Comparisons for Examining the Equality of Two Curves." Mimeo University of Toronto Economics Department.

Atkinson A.B. (1987) "On the Measurement of Poverty" Econometrica 55, 749-764.

Barrett G. and S. Donald (1999) "Consistent Tests for Stochastic Dominance" mimeo Boston University Economics Department.

Barro R.J. (1998) Determinants of Economic Growth MIT Press.

Beach C.M., R.P. Chaykowski and G.A. Slotsve (1998) Inequality and Polarization of Male earnings in the US 1968-1992 North American Journal of Economics and Finance v8 135-151

---

[12] Modification of these ideas and tests in order to accommodate population changes is a matter for further research.

Beach C.M.and G.A. Slotsve (1996) "Are we becoming 2 Societies? Income, Polarization and the Middle Class in Canada" C.D. Howe Institute.

Becker, G.S. (1981) A Treatise on the Family Cambridge Harvard University Press.

Cox D.R. (1966) "Notes on the Analysis of Mixed Frequency Distributions" British Journal of Mathematical Statistics and Psychology 19 39-47.

Davidson, R and J-Y Duclos (2000) "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. Econometrica 68 1435-1464.

Dinardo J. and T Lemieux (1997) "Diverging Male Wage Inequality in the United States and Canada 1981-1988: Do Institutions Explain the Difference?" Industrial and Labour Relations Review 50 629-651.

Doiron D. and W.G. Riddell (1994) "The Impact of Unionization on Male-Female Earnings Differences in Canada" The Journal of Human Resources 29 504-535.

Esteban, J-M. And D. Ray (1994) "On the Measurement of Polarization" Econometrica 62, 819-851.

Foster J.E. and A.F. Shorrocks (1988) "Poverty Orderings" Econometrica 56 pp.173-177.

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

Gale, D. and Shapley, L.S. (1962) "College Admissions and the Stability of Marriage" American Mathematical Monthly 69 9-15.

Gastwirth J.L., T.K. Nayak and J-L. Wang (1989) "Statistical Properties of Measures of Between-Group Income Differentials" Journal of Econometrics 42 5-19.

Gibbons J.D. and S. Chakraborti (1992) Nonparametric Statistical Inference Marcel Dekker.

Good I.J. and Gaskins R.A. (1980) "Density Estimation and Bump-Hunting by the Penalised Likelihood Method Exemplified by Scattering and Meteorite Data." Journal of the American

Statistical Association 75 42-56.

Hadar, J and W.R.Russell (1969) "Rules for Ordering Uncertain Prospects" American Economic Review 59 pp.25-34.

Hartigan, J.A. and P.M. Hartigan (1985) "The Dip Test of Unimodality" Annals of Statistics 13 70-84.

Jenkins S.P. (1996) "Recent Trends in the U.K. Income Distribution: What Happened and Why?" Oxford Review of Economic Policy 12 29-46.

Johnson, N.J., S. Kotz and N. Balakrishnan (1994) Continuous Univariate Distributions. Volume 1 2nd Edition. Wiley

Jones C.I. (1997) "On the Evolution of the World Income Distribution" Journal of Economic Perspectives 11 19-36.

Klotz J.H. (1962) "Nonparametric Tests for Scale" Annals of Mathematical Statistics 33 495-512.

Kodde, D.A. and F.C. Palm (1986) "Wald Criteria for Jointly Testing Equality and Inequality Restrictions" Econometrica 54 pp 1243-1248.

Levy, F. and R.J. Murnane (1992) "U.S. Earnings Inequality: A Review of Recent Trends and Proposed Explanations." Journal of Economic Literature 30, 1333-1381.

Levy, H. and Z. Weiner (1998) "Stochastic Dominance and Prospect Dominance with Subjective Weighting Functions" Journal of Risk and Uncertainty 16 147-163.

Mare, R. (1991) "Five Decades of Assortative Educational Mating" American Sociological Review 56 p15-32.

McFadden D. (1989) "Testing for Stochastic Dominance" in Studies in the Economics of Uncertainty: In Honor of Josef Hadar. Fomby T.K. and Seo T.K. eds. Springer. 113-134.

Mood A.M. (1954) "On The Asymptotic Efficiency of Certain Nonparametric Two Sample Tests" Annals of Mathematical Statistics 25, 514-522.

Qian Z. (1998) "Changes in Assortative Mating: The Impact of Age and Education, 1970-1990" Demography 35 p279-292.

Quah D.T. (1997) "Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs" Journal of Economic Growth 2 27-59.

Rao C.R. (1973) Linear Statistical Inference and its Applications. 2nd Edition. Wiley.

Roth, A.E. and Sotomayor, M.A.O. (1990) Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Cambridge University Press.

Silverman, B.W. (1981) "Using Kernel Density Estimates to Investigate Multimodality" Journal of the Royal Statistical Society Series B 43 97-99.

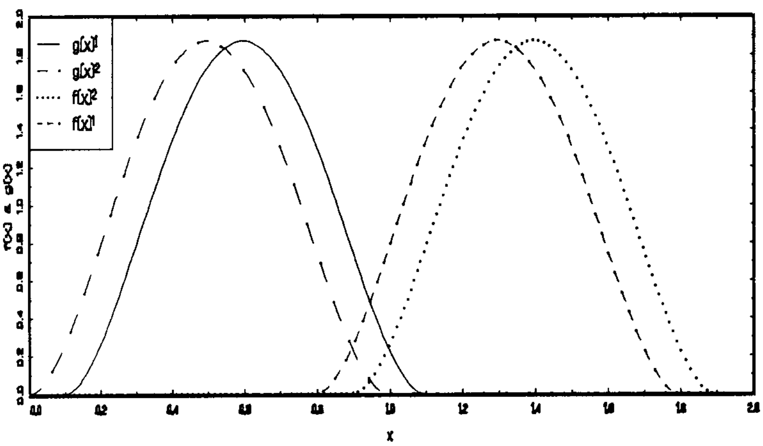
Stoline M.R. and H.A. Ury (1979) "Tables of the Studentised Maximum Modulus Distribution and an Application to Multiple Comparisons Among Means" Technometrics 21. 87-93

Weitzman, M (1970) "Measures of Overlap of Income Distributions of White and Negro Families in the U.S." Technical Paper 22 Bureau of the Census.

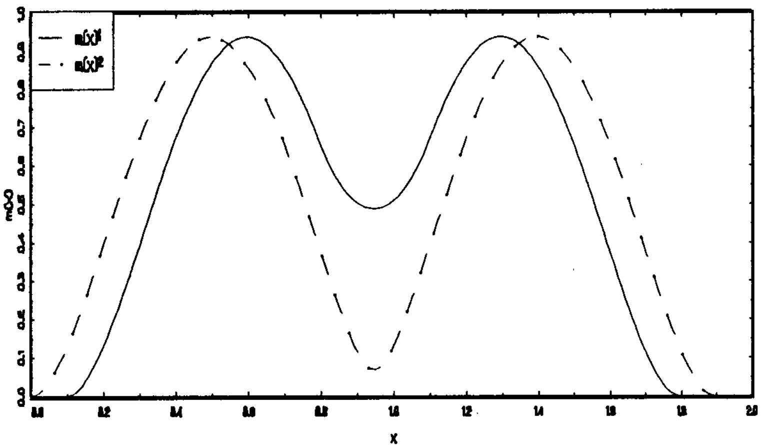
Wolak F.A. (1989) "Testing Inequality Constraints in Linear Econometric Models" Journal of Econometrics 41 205-235.

Wolfson, M.C. (1994) "When Inequalities Diverge" American Economic Review Papers and Proceedings 84, 353-358.

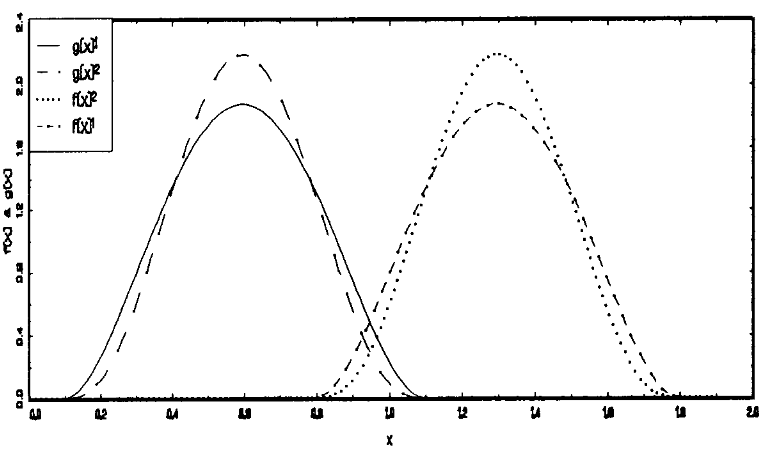
1. DIVERGENCE IN MEANS BETWEEN POPULATION POLARIZATION



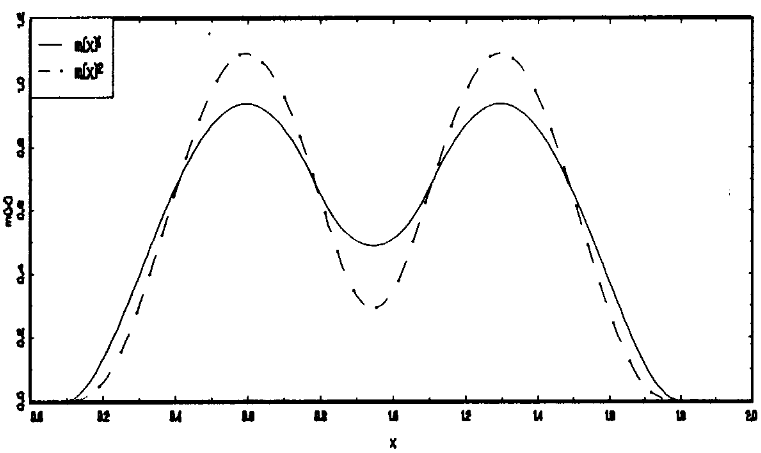
1a. DIVERGENCE IN MEANS WITHIN POPULATION POLARIZATION



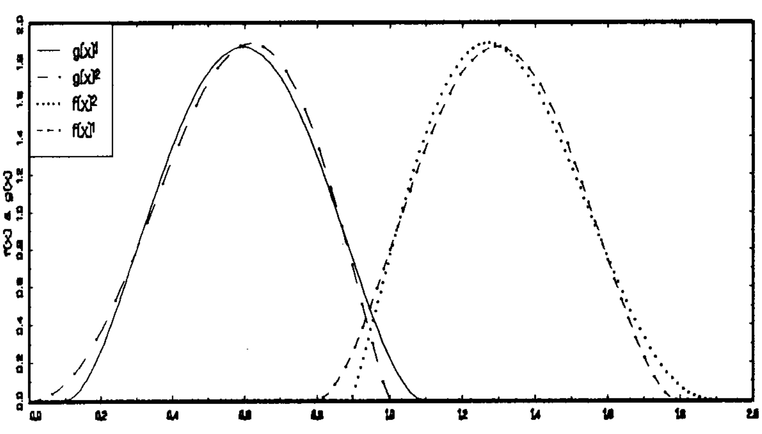
2. INCREASED CONCENTRATION BETWEEN POPULATION POLARIZATION



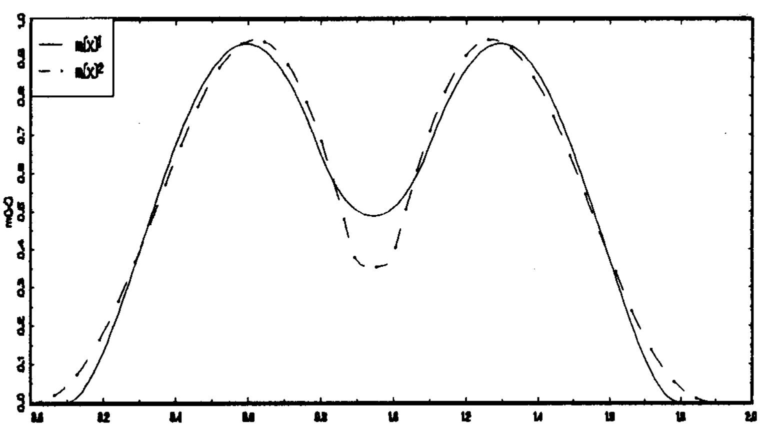
2a. INCREASED CONCENTRATION WITHIN POPULATION POLARIZATION



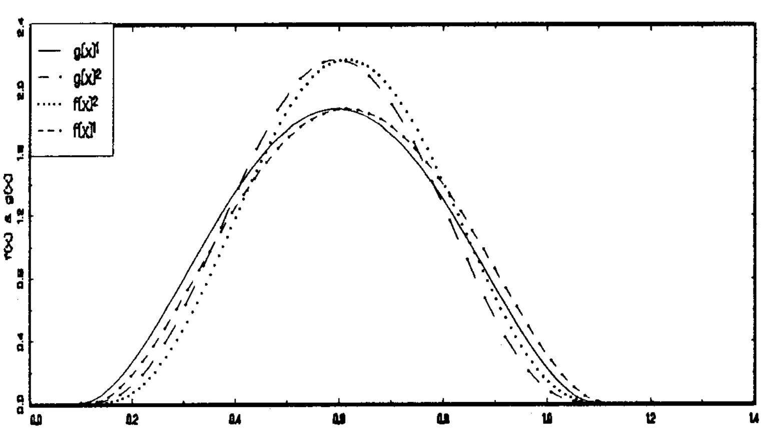
3. OPPOSITE SKEWNESS BETWEEN POPULATION POLARIZATION



3a. OPPOSITE SKEWNESS WITHIN POPULATION POLARIZATION



4. DIVERGENCE IN MEANS BETWEEN POPULATION POLARIZATION



4a. DIVERGENCE IN MEANS WITHIN POPULATION POLARIZATION