

The median as watershed¹

R Aaberge and A B Atkinson
 Statistics Norway Nuffield College, Oxford
 ESOP, University of Oslo

1. Introduction

In studies of income distribution, the median has become an increasingly important point of reference. With the fanning-out of the distribution in a number of countries, notably the United States, the mean has become a less satisfactory indicator of overall progress, and attention is turning to the median. As it was put by the Stiglitz Commission, “median consumption (income, wealth) provides a better measure of what is happening to the “typical” individual or household than average consumption (income or wealth)” (Stiglitz et al, 2009, pages 13-14 of Executive Summary). In the Europe 2020 Agenda of the European Union, the headline at risk of poverty target measures financial poverty in terms of the proportion of the population living below 60 per cent of the median. There is a burgeoning literature on the “middle class”, defined in a variety of ways, but typically in terms of a range around the median. In one approach to polarization that follows Foster and Wolfson (1992/2010), the median plays a crucial role. Progressive income transfers across the median reduce both inequality and polarization, but such transfers on one side of the median cause inequality and polarization to move in opposite directions. The median may be seen as a “watershed”: a crucial separating divide.

The aim of this paper is to bring out some of the implications of adopting the median as a dividing line for these measurement purposes, particularly with respect to the robustness of the conclusions reached. In doing so, we develop the two alternative approaches – primal and dual – applied to Lorenz curves in Aaberge (2001). In the case of poverty measurement – the first subject that we consider – the natural starting point is a poverty line defined in the income (or consumption) space, which we term the “primal” approach. However, there has been a switch in recent years to poverty lines based on quantiles, such as the EU standard set at 60 per cent of the median. This move to a “dual” approach complicates assessment of the sensitivity of the conclusions to variations in the poverty line, and has led to the search for dominance results, as we discuss in Section 2. Our focus is on “well-off” countries where poverty is a minority, rather than a majority, phenomenon. At the other end of the scale, rich people are found in all countries, but less attention has been paid to the definition of cut-offs for affluence. Definition of the “rich” may follow by default as being above the upper bound for the middle class, or they may be defined in their own right. As has been suggested by P K Sen (1988), and developed by Peichl et al (2010), the measurement of “affluence” can proceed along similar lines to the measurement of poverty. The threshold may be set, relatively, as a percentage of the median, and we can ask similar questions about the sensitivity and seek similar dominance results. These are the subject of Section 3, where we focus on societies that have a middle class in the sense that the median person is never defined as “rich”. The motivation of Foster and Wolfson’s paper “Polarization and the decline of the middle class” (1992/2010) was the sensitivity of conclusions to the

¹ We are grateful to Koen Decanq, Jean-Yves Duclos, Joan Esteban, Peter Lambert and Debraj Ray for helpful comments. They are not to be held in any way responsible for the contents of the paper.

– essentially arbitrary – definition of the middle class. They proposed “a range-free approach to measuring the middle class and polarization based on *partial* orderings” (2010, page 247). We introduce an alternative partial ordering defined in terms of a bi-polarization curve capturing the distance from the median. This approach is investigated in Section 4, where we also consider the relation to earlier concepts of dispersion and tail-heaviness and demonstrate that bi-polarization and tail-heaviness can be considered as complementary concepts of dispersion. The main conclusions are summarized in Section 5. There are two appendices covering the asymptotic theory of estimation and providing proofs of theoretical propositions.

The paper is concerned with concepts – poverty, inequality, affluence, and polarization – that are typically treated in different literatures. Our aim here is to place them within a common framework and to identify the way in which different classes of income transfers contribute to different objectives. In particular, we examine the role of transfers that preserve both the mean and the median, and the importance of distinguishing between transfers across the median and transfers on one side of the median. We begin with their implications for the measurement of poverty.

2. Measuring poverty in well-off countries

Our focus is on “well-off” countries, by which we mean countries where poverty is a minority phenomenon and where poverty is measured in relative terms, as with the 60 per cent of median income at-risk-of-poverty line, now the basis for the Europe 2020 headline target. This definition has been criticized on several grounds. Criticism 1 is the apparent arbitrariness of the choice of 60, rather than some other proportion, z (where $0 < z < 1$). It is a shortcoming that the results may depend crucially on this choice. In a single country, poverty may rise over time when measured with one value of z and fall with another. When comparing across countries, one country may have higher poverty than another country with one value of z and lower poverty with another value. A second objection is that the poverty count does not always obey the Pigou-Dalton principle of transfers. A mean-preserving transfer of income in an equalizing direction may raise recorded poverty. This can happen for two reasons. Criticism 2 is that a transfer may benefit someone well under the poverty line, but still leave them below, at the expense of a loser who was previously just above the poverty line and is brought below, increasing the poverty count. Criticism 3 is that poverty may rise because the gainer from the transfer is the median person, causing the poverty standard to rise and bring more people into the poverty net. In order to address these issues, we need to consider dominance for a range of z , and to re-consider the definition of the class of transfers.

2.1 Median-preserving across-median transfers

Criticism 3 – the fact that the mean-preserving transfer affects the poverty line – does not apply where the definition is framed, as it used to be in the EU, as a percentage of the mean. This is the first reason for introducing an alternative definition of the class of transfers, which ensures that they preserve not only the mean but also the median. The second reason is to rule out the case where the loser is brought into poverty (Criticism 2). Consider a non-negative² income variable X with cumulative distribution function F , and let F^{-1} denote the left inverse of F . Then we define

Definition 2.1. *A mean-median-preserving across-median progressive (regressive) transfer is a transfer from a person of rank t with income $F^{-1}(t)$ to a person of rank s with income $F^{-1}(s)$, where*

² The analysis can readily be extended to cover negative incomes.

$0 < s < 0.5 < t \leq 1$ ($0 < t < 0.5 < s \leq 1$), such that the transfer leaves the recipient below (above) the median and the donor above (below) the median.

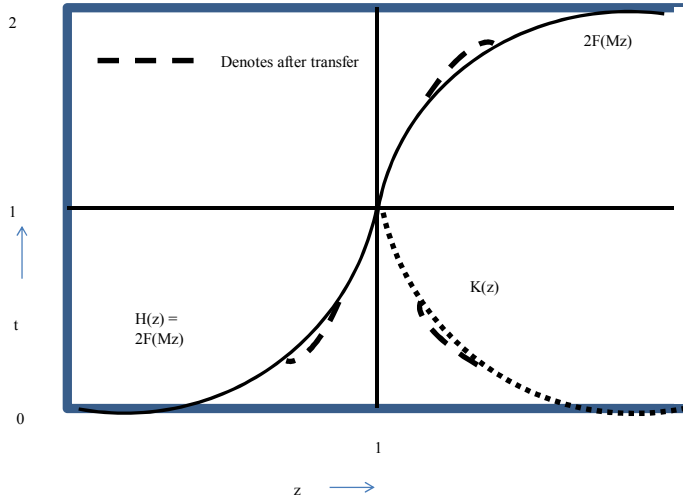
The specification that the median is unchanged implies that the progressive transfer is of sufficiently small magnitude to leave person s below the median and person t above the median (for the regressive transfer, this is not an issue).

We are assuming that, in a well-off society, the poverty line is below the median. This means that the above definition reduces the class of progressive transfers to a subset that are not subject to Criticisms 2 and 3. What about Criticism 1? Suppose for an arbitrary upper threshold, R , the associated poverty curve is defined by the conditional distribution

$\tilde{H}(z) = \Pr(X \leq zR | X \leq R) = F(zR) / F(R)$ for $0 \leq z \leq 1$.³ To avoid arbitrariness, we would like it to be the case that a mean-median preserving across-median progressive transfer is guaranteed to lower the poverty curve, and, conversely, that any lower poverty curve can be attained by a sequence of mean-median preserving across-median progressive transfers. This is only the case where $R = M$. If R were less than M , then a progressive across-median transfer could reduce $F(R)$ and hence raise the associated poverty curve. If R were greater than M , then a difference in poverty curves in the interval (M, R) requires a progressive transfer above the median.

For this reason, we take the case where $R=M$, the median. This means that we can write the poverty curve as $H(z) = 2F(Mz)$ for $0 \leq z \leq 1$. The inverse function is $H^{-1}(t) = (1/M)F^{-1}(t/2)$ for $0 \leq t \leq 1$. This is illustrated in the left hand part of Figure 1 (the right hand part is discussed later). The curve starts at the origin and ends at (1,1). Where the lowest income is strictly positive, the $H(z)$ curve follows the horizontal axis for some distance. In the case of the uniform distribution from 0 to $2M$, a useful benchmark, the $H(z)$ curve is the straight line joining the origin to the point (1,1).

Figure 1 Median-normalised poverty headcount curves and effect of across-median progressive transfer



The curve $H(z)$ is referred to as the median-normalized poverty headcount. The empirical test for one distribution being reachable by a sequence of median-preserving across-median progressive

³ Alternative poverty curves have been introduced by Atkinson (1987) and Jenkins and Lambert (1997).

transfers is that the median-normalized poverty headcount curves to be lower (or no higher) for all thresholds up to the median. Equivalently, the income at percentile t , expressed relative to the median, is higher (or no lower) at all percentiles up to 50 per cent.

Definition 2.2. A median-normalised poverty headcount curve H_1 is said to **first-degree dominate** a median-normalized poverty headcount curve H_2 if

$$H_1(z) \leq H_2(z) \text{ for all } z \in [0,1] \left(H^{-1}(t) \geq H^{-1}(t) \text{ for all } t \in [0,1] \right)$$

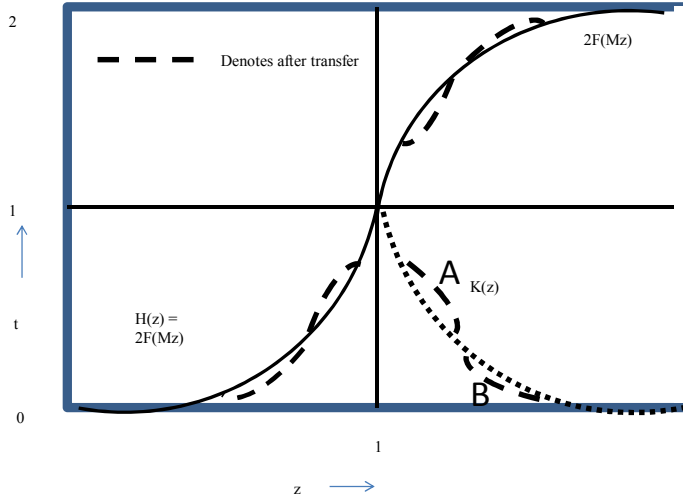
and the inequality holds strictly for some $z \in \langle 0,1 \rangle$ ($t \in \langle 0,1 \rangle$).

Since we are restricting attention to well-off countries, where poverty is a minority problem, the poverty line is never going to be in excess of the median. Put the other way round, to be sure of reducing poverty, a progressive transfer has to cross the median. A progressive transfer confined to incomes strictly above the median has no impact on poverty. A progressive transfer confined to incomes strictly below the median runs the risk of raising measured poverty for some (below median) poverty line. This is one reason why we have called the median a “watershed”. An across-median progressive transfer is illustrated by the dashed lines in Figure 1, where the $2F(Mz)$ curve extends the $H(z)$ curve for values of z greater than 1. (The $K(z)$ curve is described in Section 3.) The curve $H(z)$ after the transfer, shown by the dashed lines, is lower (or no higher), so it ensures first-degree dominance. Alternatively, viewed from the orthogonal dimension, the income associated with each t is higher (or no lower).

2.2 Second-degree dominance and transfers below the median

When median-normalised poverty headcount curves do not intersect, Theorem 2.1 provides a normative justification for using first-degree poverty (head count) curve dominance as a criterion for ranking poverty curves. Dominance is not however satisfied in the case of progressive transfers on one side of the median. As is illustrated in Figure 2, the curves intersect.

Figure 2 Effect of progressive transfers on each side of median



A transfer taking place below the median leaves the total income unchanged, which implies that the area under the H curve must be unchanged and hence that the H curves differing by such a transfer intersect. As is shown by the left hand part of Figure 2, the curve obtained after the progressive transfer is initially below and later above. Similarly as with overall inequality, we may employ second-degree poverty curve dominance defined by

Definition 2.3. A median-normalized poverty headcount curve H_1 is said to **second-degree dominate** a curve H_2 if

$$\int_0^y H_1(z) dz \leq \int_0^y H_2(z) dz \text{ for all } y \in [0, 1]$$

and the inequality holds strictly for some $y \in (0, 1)$.

In terms of Figures 1 and 2, we are integrating from the origin over z . Orthogonally, we could consider inverse stochastic dominance by integrating from the origin over t . In this case, dominance requires that the income associated with any t should be higher (the integral becomes positive), so that the condition is re-stated in terms of $\{1 - H^{-1}(t)\}$ ⁴:

$$\int_0^u (1 - H_1^{-1}(t)) dt \leq \int_0^u (1 - H_2^{-1}(t)) dt \text{ for all } u \in [0, 1].$$

In using the *inverse decumulative distribution function*, $\{1 - H^{-1}(t)\}$, we are following Yaari's dual approach for analysing income inequality (Yaari, (1987, 1988), and is therefore referred to here as the *dual* approach. The first approach corresponds to the type of income inequality analysis proposed by Atkinson (1970) and is referred to as the *primal* approach. It should be noted that the switch from the

⁴ For a proof of the equivalence between second-degree stochastic and inverse stochastic dominance see Atkinson (1970).

primal to the dual involves two changes of perspective in the bottom right hand quadrant of Figure 1: we are integrating over t rather than z , *and* we are taking the complement of the function. This second change is parallel to the concept of “downwards Lorenz dominance” in Aaberge (2009). In the next section we show that second-degree poverty dominance can be given a normative justification in terms of application of a combination of below median progressive transfers and median-preserving across-median progressive transfers.

2.3 Towards a complete ranking: degree and weighting

A social decision-maker who employs the second-degree dominance of two intersecting poverty curves pays more attention to the poorest persons than to the less extreme poor. The degree to which more weight is attached does not matter providing that the second-degree dominance condition is satisfied. Where, however, there are transfers below the median in both directions, a complete ranking can only be attained if a specific weighting function is introduced. There are in fact two steps. We observe that a person has income Mz and that the person ranks $F(Mz)$ in the distribution. First, we ask whether the person is of concern and to what degree. Above we have followed two approaches. One, the primal approach, asks about the person’s income: that z is less than or equal to 1, with $H(z)$ being a measure of degree. The alternative, dual, approach asks about the person’s rank: that F is less than or equal to $1/2$, with $(1-z)$ being a measure of the degree.

The second step introduces weights. In general, these weights could depend on both elements: z and H . However, the measures typically used simplify by assuming that the poverty index is linear in degree: H in the primal case and $(1-z)$ in the dual case. Linearity is a strong assumption, and assumes a specific form of independence – see below. In what follows we assume such linearity, although we should note that it would be interesting to explore the implications of relaxing this assumption.⁵ Following the primal approach, linearity implies that the poverty head count curves can be ranked by the following criteria,

$$(2.1) \quad \Psi_a = \int_0^1 a(z)H(z)dz = 2 \int_0^1 a(z)F(Mz)dz ,$$

where $a(z)$ is a positive non-increasing weighting function, and lower Ψ_a means lower poverty. Linearity implies that $a(z)$ does not depend on $H(z)$. The severity of a particular proportion of the population being below the poverty line depends solely on the distance they are from the poverty line. Note that members of the family of poverty measures defined by (2.1) can be interpreted as weighted averages of poverty headcounts where the poverty threshold varies from 0 to M , and can thus be considered as threshold-free measures of poverty. Note also that the weight attached to z can be interpreted as the marginal valuation of income, since⁶

$$(2.2) \quad \int_0^1 m(z)dH(z) = 2 \int_0^1 m(z)dF(Mz) = m(1) - 2 \int_0^1 m'(z)F(Mz)dz$$

and second-order poverty head count curve dominance requires that $m'(z) = a(z)$ is non-increasing. In other words, we require that $m(z)$ is a weakly concave function.

⁵ See Aaberge and Mogstad (2010), who build on the work of Green and Jullien (1988) on the theory of choice under uncertainty.

⁶ We do not discuss here the measurement of overall inequality, but we may note that the corresponding primal measure involves comparing the integral of F weighted by the marginal valuation of income and the dual measure involves the integral of income weighted by a function of F (e.g. the Gini coefficient).

In contrast, with the dual approach the linearity assumption leads to the head count curves being ranked by the following criteria,

$$(2.3) \quad \Pi_p = \int_0^1 p(t)(1 - H^{-1}(t))dt = \frac{2}{M} \int_0^{\frac{1}{2}} p(2t)(M - F^{-1}(t))dt,$$

where $p(t)$ is a positive non-increasing weighting function, and lower Π_p means lower poverty. In this case, linearity means that the severity of a particular income gap is weighted solely according to the proportion of the population at or below that level. As is demonstrated by Theorem 2.2 below, the conditions of decreasing $p(t)$ and $a(z)$ make sure that Π_p and Ψ_a rank poverty curves consistently with second-degree poverty curve dominance. Moreover, we may note that the decreasing assumption rules out the headcount measure. With the headcount, the marginal value of an extra unit of income is zero at incomes strictly below the poverty line, becomes positive at it takes a person over the poverty threshold, and then falls again to zero. In this way, the criticism 2 is addressed (Atkinson, 1987). To ensure that Ψ_a and Π_p have the unit interval as their range we restrict attention to weighting functions a and p such that $a(1) = 0$ and $p(1) = 0$.

The two versions of the linearity assumption may, as in the literature on uncertainty, be derived from underlying independence axioms governing the ordering \succeq on \mathbf{H} , which is assumed to be continuous, transitive and complete and to rank $H_1 \succeq H_2$ if $H_1(z) \leq H_2(z)$ for all $z \in [0, 1]$,

Axiom (Independence). Let H_1 , H_2 and H_3 be members of \mathbf{H} and let $\alpha \in [0, 1]$. Then $H_1 \succeq H_2$ implies $\alpha H_1 + (1 - \alpha)H_3 \succeq \alpha H_2 + (1 - \alpha)H_3$.

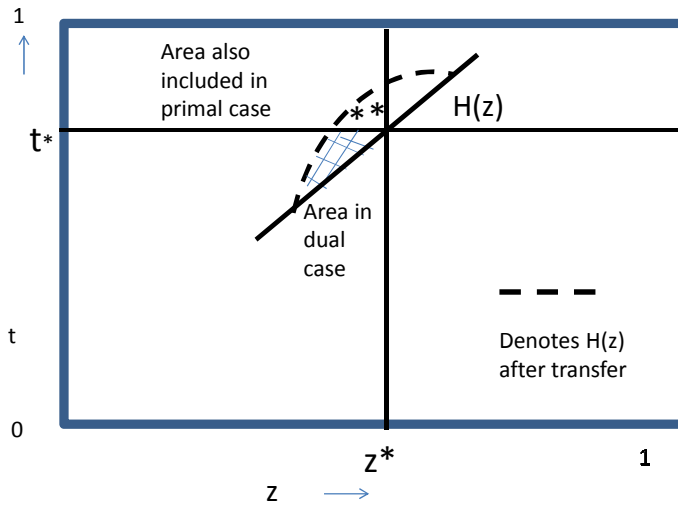
Axiom (Dual Independence). Let H_1 , H_2 and H_3 be members of \mathbf{H} and let $\alpha \in [0, 1]$. Then $H_1 \succeq H_2$ implies $(\alpha H_1^{-1} + (1 - \alpha)H_3^{-1})^{-1} \succeq (\alpha H_2^{-1} + (1 - \alpha)H_3^{-1})^{-1}$.

The axioms require that the ordering is invariant with respect to certain changes in the head count curves being compared. It is these axioms that give the preferences of the planner an empirical content. If H_1 is weakly preferred to H_2 , then the Independence Axiom of expected utility theory states that any mixture on H_1 is weakly preferred to the corresponding mixture on H_2 . The intuition is that identical mixing interventions on the head count curves being compared do not affect the ranking of head count curves; the ranking depends solely on how the differences between the mixed head count curves are judged. Thus, the axiom requires the ordering relation \succeq to be invariant with respect to aggregation of sub-populations across median relative income ratios. It means that if there is more poverty in a sub-group (in H_2), then, other things equal, there is more poverty overall. The Dual Independence axiom postulates a similar invariance property on the inverse head count curves, or the income gaps. It says that, if we consider a decomposition by income source, then dominance with regard to one source implies, other things equal, overall dominance. The essential difference between the two axioms is that the Independence Axiom deals with the relationship between given income ratios and weighted averages of corresponding population proportions, while the Dual Independence Axiom deals with the relationship between given population proportions and weighted averages of corresponding income ratios relative to the median.

The two approaches can lead to different conclusions. As an illustration, consider the case where the weighting functions $a(z)$ and $p(t)$ take on only the values 1 or 0, switching from 1 to 0 at some point. In the primal case, all concern is focused on people with incomes below a specified level; in the second case, all concern is focused on the bottom t per cent. Suppose that the switching points are z^* and $t^* = H(z^*)$, and that there is a transfer from people above z^* (but below the median) to

those below z^* , as shown in Figure 3. On the first approach, the difference in the weighted integral is the hatched area, counting the proportion raised to z^* . On the second approach, the difference also includes the starred area, counting the full income gains to the bottom t^* per cent. A person concerned to target transfers may view payments beyond those necessary to bring people to z^* as a sign of “inefficiency”. If so, they would follow the first approach. On the other hand, those concerned with the circumstances of the bottom t^* would want to count the full gain to their incomes, and hence follow the second approach. Where the switching point occurs in a range containing donors, rather than recipients, as in Figure 4, then the positions are reversed. The dual approach considers only the hatched area, whereas the primal approach also includes the starred area, composed of losses to people outside the bottom t^* , and hence regarded as irrelevant by those focused on this group.

Figure 4 Effect of weighting under primal and dual approaches: donors of transfer



As suggested above, the criterion of second-degree poverty curve dominance can be used to justify the conditions of decreasing weighting functions $p(t)$ and $a(z)$. Theorem 2.1 provides a normative justification of these conditions. Next, let \mathbf{H} denote the family of H curves.

Theorem 2.1. Let H_1 and H_2 be members of \mathbf{H} . Then the following statements are equivalent,

- (i) H_1 second-degree dominates H_2
- (ii) H_1 can be attained from H_2 by application of a combination of below median progressive transfers and median-preserving across-median progressive transfers.
- (iii) $\Psi_a(H_1) < \Psi_a(H_2)$ for all positive decreasing a such that $a(1) = 0$.
- (iv) $\Pi_p(H_1) \leq \Pi_p(H_2)$ for all positive decreasing p such that $p(1) = 0$.

(Proof in Appendix 2).

2.4 Specific poverty measures

Selecting a particular functional form for the weighting functions $a(z)$ and $p(t)$ leads to specific poverty measures. A number of the widely-employed measures are obtained by considering members of the classes:

$a_k(z) = k(1-z)^{k-1}$ and $p_k(t) = k(1-t)^{k-1}$ where $k \geq 1$, which form the following two alternative families of poverty measures,

$$(2.4) \quad \Psi_a = \Psi_k \equiv k \int_0^1 (1-z)^{k-1} H(z) dz = \frac{2k}{M^k} \int_0^M (M-x)^{k-1} F(x) dx, \quad k \geq 1$$

and

$$(2.5) \quad \Pi_p = \Pi_k \equiv k \int_0^1 (1-t)^{k-1} (1-H^{-1}(t)) dt = \frac{2k}{M} \int_0^{\frac{1}{2}} (1-2t)^{k-1} (M-F^{-1}(t)) dt, \quad k \geq 1.$$

Where $k = 1$, this yields

$$(2.6) \quad \Psi_1 = 2 \int_0^1 F(Mz) dz = \frac{2}{M} \int_0^{\frac{1}{2}} t dF^{-1}(t) = \frac{M - E(X|X \leq M)}{M}.$$

This is the poverty gap in the case where the median is the threshold, expressed relative to the threshold. The poverty gap can only be reduced by progressive across-median transfers. In this case, the same value is obtained for Π_1 . From the proof of Theorem 5 in Aaberge (2001), it may be seen that the poverty gap is the only measure of poverty that satisfies both the primal and the dual independence axioms. Thus, the primal and the dual independence axioms provide together with the conditions of transitivity, completeness, continuity and first-degree dominance a complete axiomatic characterization of the poverty gap.

The $k = 1$ measure is the same for both primal and dual approaches, but they depart for values of k in excess of 1. We can in fact see that the primal and dual approaches respectively generate different well-known poverty measures. On the primal approach, if we take $k = 2$ and let $a(z) = 2(1-z)$, then

$$(2.7) \quad \Psi_2 = \left(1 - \frac{E(X|X \leq M)}{M}\right)^2 + \frac{\text{var}(X|X \leq M)}{M^2} = \Psi_1^2 + \frac{\text{var}(X|X \leq M)}{M^2}.$$

The minimum value 0 is attained when the poorest 50 per cent of the population has equal incomes (equal to M). In our “well-off” society, there is indeed then no poverty. Expression (2.7) shows that Ψ_2 decreases with increasing average income and decreasing income dispersion for the poorest 50 per cent of the population. In contrast, on the dual approach, if we let $p=2(1-t)$, then we obtain a Gini version:

$$(2.8) \quad \Pi_2 = \frac{M - E(X|X \leq M)}{M} + \frac{E(X|X \leq M)}{M} G_l = \Pi_1 + (1 - \Pi_1) G_l,$$

where G_l denotes the lower tail Gini coefficient, i.e. the Gini coefficient of the conditional distribution of X given that $X \leq M$. This leads to a measure closely related to the poverty measure introduced by Sen (1976) and coincides with the modified Sen measure proposed by Shorrocks (1995) when the poverty line is equal to M . Moreover, it can be demonstrated that replacing H with \tilde{H} in (2.6) for

$p(t) = 2(1-t)$ actually will lead to the poverty line dependent measure introduced by Shorrocks (1995).

What happens if we take k larger than 2? The poverty index may be written in terms of the integral of the weighting function (see equation (2.2)). In the primal case, this is $(1-z)^k$. The index may be seen therefore as the analogue of the index proposed by Foster, Greer and Thorbecke (1984), referred to as the FGT index, with progressively higher values of k attaching more and more weight to the largest poverty gaps. In the same way, with the dual approach, the integral of the weights is $(1-t)^k$. Note that the most poverty averse Π_p -measure is obtained as k approaches ∞ . In this case the poverty measure is defined by

$$(2.9) \quad \Pi_{\infty} = \frac{M - F^{-1}(0+)}{M},$$

where $F^{-1}(0+)$ is the lowest income.

An interesting question is whether Π_k for $k > 2$ is related to summary measures of inequality in a similar way as Π_2 is related to the lower tail Gini coefficient. By noting that the lower tail version of the extended Gini family of inequality measures (Donaldson and Weymark, 1980) can be expressed as follows

$$(2.10) \quad G_{l,k} = 1 - \frac{2k}{\mu_l} \int_0^{\frac{1}{2}} (1-2t)^{k-1} F^{-1}(t) dt, \quad k \geq 2,$$

where $\mu_l = E(X | X \leq F^{-1}(u))$, we get the following alternative expression for Π_k by inserting (2.10) into (2.5),

$$(2.11) \quad \Pi_k = 1 - \frac{\mu_l}{M} (1 - G_{l,k}) = \Pi_1 + (1 - \Pi_1) G_{l,k}, \quad k \geq 1.$$

Similarly, it can be shown that Ψ_k is determined by the k first moments of the conditional distribution H of X given that $X \leq M$,

$$(2.12) \quad \Psi_k = \frac{E[(M-X)^k | X \leq M]}{M^k} = \sum_{i=0}^k (-1)^i \binom{k}{i} \frac{E(X^i | X \leq M)}{M^i}, \quad k \geq 1.$$

When k increases, Ψ_k becomes more sensitive to income changes that concern the poorest people.

In the formulation above, we allowed for the possibility that the poverty line could take any value up to the median: they are “threshold-free” in the sense that they apply for all poverty lines (up to the watershed). Where the threshold is known as a fraction of the median, then this replaces 1 in the weighting function and in the limit of integration. In this way, we can see that there is a duality between the class of FGT indices of poverty and the class based on the Sen index. The Sen index itself corresponds to the squared variance (coefficient of variation) version of the FGT index, and generalisations of the Sen index correspond to versions of the FGT index with k greater than 2. The duality, and the relation to the two independence axioms, illuminates further aspects not discussed here, such as the link with sub-group consistency and decomposability. More generally, the framework

proposed for considering incomes below the watershed allows us to see in a unified way the different steps and measures involved in the measurement of poverty. We now turn to the other end of the scale.

3. Measuring affluence

Most attention focuses on the lower part of the income distribution, but a number of studies have sought to apply similar techniques to the study of “affluence”. Rather than considering the top 1 per cent, say, these studies have defined a cut-off above which people can be classified as “rich”, thus allowing the proportion of rich people to vary. There will always be a top 1 per cent, but a society may limit the number of people with incomes above the “affluence” cut-off, these incomes being adjudged beyond the limits of affluence, or “excessive”. For example, lines of “affluence” have been defined as percentages of the median. Peichl, Schaefer and Scheicher (2010, page 608) take the richness line to be twice the median, describing it as “arbitrary but common practice”, whereas Brzezinski (2010) also considers lines equal to three and four times the median. There has been less discussion of the underlying theory, but an important exception is the article by P K Sen, where he presents, “side by side, some of the poverty indexes [and a] proper motivation and formulation of parallel indexes of affluence” (Sen, 1988, page 66).

3.1 Headcounts of affluence

Pursuing this parallel, we let the head count curve of affluence \tilde{K} be defined by

$$(3.1) \quad \tilde{K}(z) = \Pr(X \geq Rz | X \geq R) = \frac{1 - F(Rz)}{1 - F(R)}, \quad z \geq 1,$$

where R is a lower threshold. For a given threshold R , $\tilde{K}(z)$ shows the proportion of the (adult) population classified as richer than $100z$ per cent of R . The head count curve \tilde{K} can be used as a basis for ranking distributions with regard to affluence. Thus, for a given threshold R the higher head count curve exhibits highest affluence. However, parallel criticisms arise to those in the case of poverty measurement. Criticism 1 is the apparent arbitrariness of the choice of twice, or some other multiple, of the median. Again, the results may depend crucially on this choice. When comparing across countries, one country may have more rich people if the threshold is 200 per cent of the median, but fewer if the threshold is 300 per cent of the median. Again, the second objection is that the affluence count does not always obey the Pigou-Dalton principle of transfers. Criticism 2 is that a progressive transfer from someone well above the affluence threshold may benefit someone below the cut-off by enough to raise them above the cut-off, thus increasing the proportion of rich people. Criticism 3 is that, where the cut-off is a multiple of the median, measured affluence may rise because the loser from the transfer is the median person, causing the affluence standard to rise and bring more people into category of rich.

As before, we seek to meet these objections by restricting the class of transfers. This means that the empirical test for one distribution being reachable by a sequence of mean-median-preserving across-median progressive transfers is that the headcount curve of affluence to be lower (or no higher) for all thresholds above the median. Since we are restricting attention to countries where there is a middle class in the sense that the median person is never defined to be rich, a progressive transfer confined to incomes strictly below the median has no impact on measured affluence. A progressive transfer confined to incomes strictly above the median runs the risk of raising measured affluence for some (above median) cut-off.

The headcount curve of affluence may - following the approach adopted to the measurement of poverty - be normalised by the median:

$$K(z) = 2(1 - F(Mz)), 1 \leq z < \infty, \quad K^{-1}(t) = \frac{1}{M} F^{-1}\left(1 - \frac{t}{2}\right), 0 \leq t \leq 1$$

where K^{-1} is the left inverse of K , and let \mathbf{K} denote the family of median head count curves where $R=M$. As before, we can set out dominance conditions:

Definition 3.1. A median-normalised affluence headcount curve K_1 is said to **first-degree dominate** a median-normalised affluence headcount curve K_2 if

$$K_1(z) \leq K_2(z) \text{ for all } z \in [0,1] \left(K_1^{-1}(t) \leq K_2^{-1}(t) \text{ for all } t \in [0,1] \right)$$

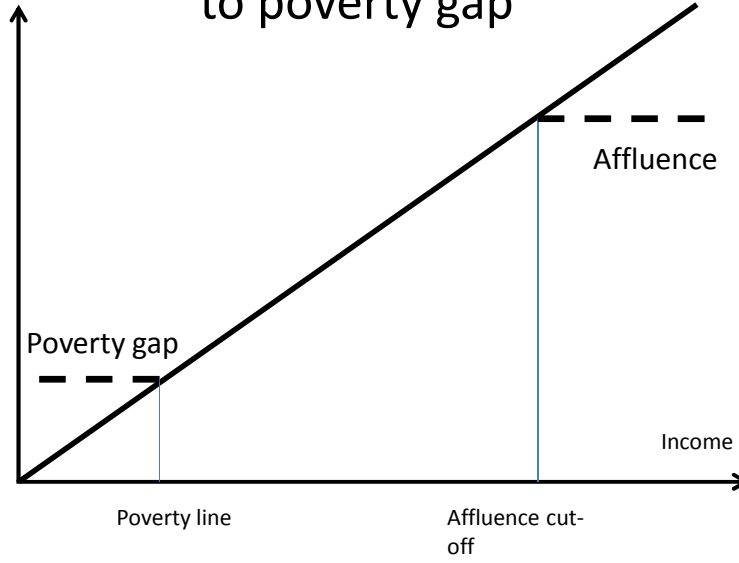
and the inequality holds strictly for some $z \in (0,1)$ ($t \in (0,1)$).

The headcount curves are illustrated in the right hand part of Figures 1 and 2. This construction is due to Foster and Wolfson (1992/2010), where we have simply turned their Figure 9 upside down. Figure 1 shows the impact of a progressive across-median transfer, and it may be seen that the after-transfer K curve dominates. On the other hand, Figure 2 shows that a progressive transfer above the median leads to intersecting K curves. A transfer taking place above the median leaves the total income unchanged, which implies that the area under the K curve must be unchanged and hence that the K curves differing by such a transfer intersect. This brings us to higher-degree dominance.

3.2 Second-degree dominance

As with poverty measurement, where the $K(z)$ curves intersect, we may extend the ranking by making stronger assumptions. Following the parallel with poverty measures, the poverty gap has a natural analogue: the “excess” income of the rich over and above the affluence cut-off – see Figure 5. For a cut-off of z times the median, this generates a measure which is the income share of the rich group minus Mz/μ times the proportion rich, where μ is the mean income. So that, with an income share of the top 1 per cent of 12 per cent, and a cut-off of 4 times the mean, the excess income is 8 per cent.

Figure 5 Measure of affluence parallel to poverty gap



The excess income indicator may be seen as attaching a value of 1 to a marginal unit of income below Mz , and of zero to a marginal unit of income above Mz . This is a non-increasing function. It therefore falls (or is unchanged) if there is a progressive transfer within the top half. However, in contrast to the poverty curve, $H(z)$, this progressive transfer has the effect of raising $K(z)$ at the lower end and lowering it further up – see Figure 2. This means that we cannot simply integrate over $K(z)$ from $z = 1$ upwards and require that the cumulative difference be negative (or zero). In order for the difference in the integral to be negative, we have to integrate downwards, following the approach of Aaberge (2009), applied there to Lorenz curves: i.e. to integrate over the range from u to infinity, and require that this be negative for all u greater than equal to 1. In this way, the area B in Figure 2 is said to be more important than area A.

A transfer taking place above the median leaves the total income unchanged, which implies that the area under the K curve must be unchanged and hence that the K curves differing by such a transfer must intersect. Similarly as with poverty, we may employ second-degree affluence curve dominance defined by

Definition 3.2. A median-normalised affluence headcount curve K_1 is said to **second-degree dominate** a median-normalised affluence headcount curve K_2 if

$$\int_0^u K_1^{-1}(t)dt \leq \int_0^u K_2^{-1}(t)dt \text{ for all } u \in [0,1]$$

and the inequality holds strictly for some $u \in [0,1]$.

Note that the dominance condition of Definition 3.2 can be considered as second-degree downward dominance of the income distribution F (relative to its median), which follows from the fact that

$$\int_0^u K^{-1}(t)dt = \frac{2}{M} \int_{1-\frac{u}{2}}^1 F^{-1}(t)dt \text{ for all } u \in [0,1].$$

Thus, aggregation starts from the highest incomes whereas second-degree poverty dominance aggregates incomes from below and starts with the lowest incomes. As is demonstrated by Theorem 3.1 below, a social decision-maker who employs the second-degree dominance of two intersecting affluence curves pays more attention to a transfer from the richest person than to transfers from the less extreme rich.

3.3 Towards a complete ranking

A social decision-maker who employs the second-degree dominance of two intersecting affluence curves pays more attention to the persons nearest to the cut-off. The degree to which more weight is attached does not matter where the second-degree dominance condition is satisfied. Where, however, there are transfers above the median in both directions, a complete ranking can only be attained if a specific weighting function is introduced. As before, this can be approached from the standpoint of either the primal or the dual.

From the primal standpoint, the independence axiom (replacing H by K), implies that the head count curves should be ranked by the following criteria,

$$(3.2) \quad A_b = \int_1^{\infty} b(z)K(z)dz = 2 \int_1^{\infty} b(z)(1 - F(Mz))dz = \frac{2}{M} \int_M^{\infty} b\left(\frac{x}{M}\right)(1 - F(x))dx,$$

where $b(z)$ is a positive weighting function. In contrast, with the dual approach, the dual independence axiom (with H replaced by K) implies that the head count curves K should be ranked by the following criteria,

$$(3.3) \quad \Gamma_q = \int_0^1 q(t)(K^{-1}(1-t) - 1)dt - 1 = \frac{2}{M} \int_{\frac{1}{2}}^1 q(2t-1)(F^{-1}(t) - M)dt,$$

where $q(t)$ is a positive weighting function.

Similarly as for the poverty measures we impose the normalizing conditions $b(1) = 0$ and $q(0) = 0$ on the weighting functions b and q . To impose further restrictions on the weighting functions q and b it appears attractive to explore the relationship between second-degree affluence curve dominance and the families Γ_q and A_b of affluence measures. The following characterization result shows that it is necessary to restrict the weighting functions q and b to be increasing to ensure equivalence between second-degree affluence curve dominance and Γ_p - and Γ_b -measures as decision criteria (in the case of poverty measures, the weighting functions were decreasing).

Theorem 3.1. Let K_1 and K_2 be members of \mathbf{K} . Then the following statements are equivalent,

- (i) K_1 second-degree dominates K_2 .
- (ii) K_1 can be attained from K_2 by application of a combination of above median progressive transfers and mean-median-preserving across-median progressive transfers.
- (iii) $\Gamma_q(K_1) < \Gamma_q(K_2)$ for all positive increasing q such that $q(0) = 0$.
- (iv) $A_b(K_1) \leq A_b(K_2)$ for all positive increasing b such that $b(1) = 0$.

(Proof in Appendix 2).

As before, we can consider the specific class of weighting functions $b(z) = k(z-1)^{k-1}$, which leads to the following family of affluence measures,

(3.4)

$$A_b = A_k \equiv k \int_1^\infty (z-1)^{k-1} K(z) dz = \frac{2k}{M^k} \int_M^\infty (x-M)^{k-1} (1-F(x)) dx = \frac{E\left((X-M)^k | X \geq M\right)}{M^k}.$$

The latter term is obtained by using integration by parts.

Where $k = 1$

$$(3.5) \quad A_1 = \frac{2}{M} \int_M^\infty (1-F(x)) dx = \frac{E(X|X \geq M) - M}{M}.$$

When $F \sim \text{unif}(0, c)$ then $A_1 = \frac{1}{2}$. Note that A_1 can be considered as a threshold-free head count measure of affluence as well as a measure of the affluence gap. Where $k=2$,

$$(3.6) \quad A_2 = A_1^2 + \frac{\text{var}(X|X \geq M)}{M^2}$$

which shows that A_2 decreases with decreasing average income and decreasing income dispersion for the richest 50 per cent of the population. The minimum value 0 is attained when the richest 50 per cent of the population all have incomes equal to M .

As for the dual poverty measures, we can consider the analogous specific class of positive increasing weighting functions $q(t) = kt^{k-1}$ for $k > 1$, which forms the following dual family of affluence measures,

$$(3.7) \quad \Gamma_k = k \int_0^1 t^{k-1} (K^{-1}(1-t) - 1) dt - 1 = \frac{2k}{M} \int_{\frac{1}{2}}^1 (2t-1)^{k-1} (F^{-1}(t) - M) dt.$$

Where $k=1$, the measures coincides with the primal measure

$$(3.8) \quad \Gamma_1 = \frac{E(X|X \geq M) - M}{M} = A_1.$$

For higher values of k , the measures differ, Where $k=2$, a Gini version of Γ_q is obtained:

$$(3.9) \quad \Gamma_2 = \frac{E(X|X \geq M) - M}{M} + \frac{E(X|X \geq M)}{M} G_u = \Gamma_1 + (1 + \Gamma_1) G_u,$$

where G_u denotes the Gini coefficient of the conditional distribution of X given that $X \geq M$. Note that the most affluence-averse Γ_p -measure is obtained as k approaches ∞ . In this case the affluence measure is defined by

$$(3.10) \quad \Gamma_{\infty} = \frac{F^{-1}(1) - M}{M},$$

where $F^{-1}(1)$ is the highest income. Similarly, as for the analogous family Π_k of poverty measures, we find that Γ_k can be expressed in terms of measures of inequality,

$$(3.11) \quad \Gamma_k = \left[(k-1)D_{u,k} + 1 \right] \frac{\mu_u}{M} - 1 = \Gamma_1 + (k-1)(1 + \Gamma_1)D_{u,k}, \quad k \geq 2,$$

where

$$D_{u,k} = \frac{2k}{(k-1)\mu_u} \int_{\frac{1}{2}}^1 (2t-1)^{k-1} F^{-1}(t) dt - \frac{1}{k-1}, \quad k \geq 2$$

defines the upper tail version of the Lorenz family of inequality measures (Aaberge, 2000) and $\mu_u = E(X|X \geq M)$. For increasing k , $D_{u,k}$ increases its weight on progressive transfers the further up in the income distribution they take place.

On the primal approach, we get by using a Taylor expansion that

$$(3.12) \quad A_k = \frac{2}{M^k} \int_M^{\infty} (x-M)^k dF(x) = \sum_{i=0}^k (-1)^i \binom{k}{i} \frac{E(X^i | X \geq M)}{M^i},$$

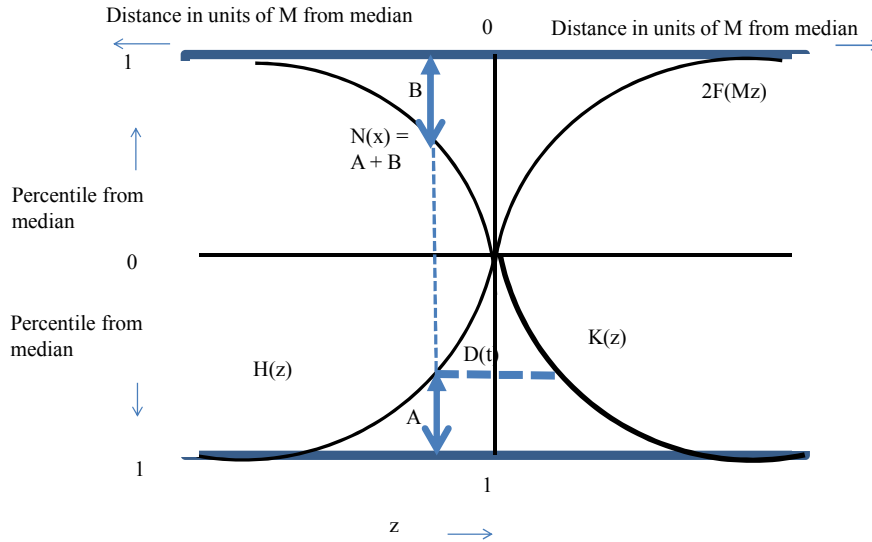
which demonstrates that A_k is determined by the k first moments of the conditional distribution of X given that $X \geq M$. When k increases A_k becomes more sensitive to changes that concern the most affluent people.

In the next section, we show how dispersion in the distribution as a whole can be related to the pairs of affluence and poverty measures: the poverty measures Ψ_k and the affluence measure A_k in the primal case, and the poverty measure Π_k and the affluence measure Γ_k in the dual case.

4. Dispersion, bi-polarization and tail-heaviness

We now bring together the two halves of our discussion and consider the distribution as a whole. This is necessary to address the much-discussed issue of the decline of the middle class, in the form of a shift in weight away from the median towards the tails of the distribution. We refer to this as “bi-polarization”, to distinguish it from other concepts of polarization, notably those pioneered by Esteban and Ray (1994, 1999 and 2012) and Duclos, Esteban and Ray (2004). It also allows us to consider the concepts of *tail-heaviness* and *dispersion* introduced in the earlier statistical literature. Combining the two curves in Figure 1 (and 2), $H(z)$ and $K(z)$, does in fact suggest a natural way to measure the extent of dispersion, or, conversely, the extent of concentration: the distance in terms of income (defined relative to the median) between percentiles equi-distant from the median – see the distance $D(t)$ in Figure 6. It is with the concept of dispersion that we begin.

Figure 6 Measures of dispersion



4.1. Dispersion

A general definition of dispersion is given by Bickel and Lehmann (1979, page 34) as follows: the distribution F is less dispersed than the distribution G if for all $0 < u < v < 1$, $F^{-1}(v) - F^{-1}(u)$ is less than (or equal to) $G^{-1}(v) - G^{-1}(u)$ ⁷. In other words, it requires that two quantiles of G are at least as far apart as the corresponding quantiles of F . Here we apply a weaker version where $u = (1-t)/2$ and $v = (1+t)/2$. In other words, we use the following curve, denoted the dispersion curve,

$$(4.1) \quad D(t) = \frac{1}{M} \left(F^{-1}\left(\frac{1+t}{2}\right) - F^{-1}\left(\frac{1-t}{2}\right) \right), \quad t \in [0,1],$$

⁷ Note that Doksum (1969) introduced another form of the Bickel-Lehmann condition as a tail-ordering.

where $D(0) = 0$, since we are then looking at $F^{-1}(1/2)$ in both cases. As t approaches 1, the distance becomes $D(1) = (F^{-1}(1) - F^{-1}(0))/M$. This is illustrated in Figure 6, where it should be noted that $D(t)$ is measured at $(1-t)$ (i.e. below the median), the value for $(1+t)$ being obtained from the reflection of $2F(Mz)$ in the form of $K(z)$.

Since dispersion $D(t)$ is defined in terms of t , we refer to this as the dual approach. In Figure 6, we also show an alternative primal distance measure, $N(z)$, where z is defined in terms of cumulative frequencies at the same distance from the median: i.e. $F(M(1+z))$ and $F(M(1-z))$, where $F(x)$ is defined as zero for all negative x values. As in earlier sections, the move from dual to primal involves two steps: the move from ranks to incomes, and taking the complement. The measure $N(z)$ is therefore defined as.

$$(4.2) \quad N(z) = 2[Pr(X < M(1-z)) + Pr(X > M(1+z))] = 2[F(M(1-z)) + 1 - F(M(1+z))], z \geq 0.$$

As with the dual, Figure 6 shows $N(z)$ measured at $(1-z)$, to the left of the median, with the value for $(1+z)$ being obtained from the reflection of $2F(Mz)$. With the primal approach, rather than comparing, say, the distance between the upper and lower quartiles, we are asking what proportion are above, say, 125 per cent of the median or below 75 per cent of the median. When $z = 0$, this proportion is 100 per cent; as z gets large, the proportion goes to zero. In the dual case, if $D_1(t) \leq D_2(t)$ for all $t \in [0,1]$ then we say that D_1 exhibits dispersion dominance of first-degree; i.e. the distribution F_1 (associated with D_1) exhibits less dispersion than the distribution F_2 (associated with D_2). In this case it is clear that D_1 (and the corresponding distribution function F_1) can be obtained from D_2 (and the corresponding distribution function F_2) by employing progressive transfers below, above as well as across the median (see Definitions 2.1, 2.3 and 3.2)).

To deal with situations where the dispersion curves $D(t)$ intersect, which normally will be the case in empirical applications, a weaker criterion than first-degree dispersion dominance is called for. Two alternative dominance criteria emerge as natural candidates; one that aggregates the dispersion curve from below (second-degree *upward* dispersion dominance) and the other that aggregates the dispersion curve from above (second-degree *downward* dispersion dominance). It should be noted that first-degree dispersion dominance implies second-degree upward as well as downward dispersion dominance. However, the transfer sensitivity of these criteria differ in the sense that second-degree *upward* dispersion dominance places more emphasis on transfers occurring in the central part around the median rather than in the lower and upper part of the income distribution, whereas second-degree *downward* dispersion dominance is most sensitive to transfers that occur at the tails of the income distribution.

Definition 4.1a. A dispersion curve D_1 is said to second-degree upward dominate a dispersion curve D_2 if

$$\int_0^u D_1(t) dt \leq \int_0^u D_2(t) dt \text{ for all } u \in [0,1]$$

and the inequality holds strictly for some $u \in (0,1)$.

Definition 4.1b. A dispersion curve D_1 is said to second-degree downward dominate a dispersion curve D_2 if

$$\int_u^1 D_1(t)dt \leq \int_u^1 D_2(t)dt \text{ for all } u \in [0,1]$$

and the inequality holds strictly for some $u \in \langle 0,1 \rangle$.

The normative content of second-degree upward and downward dispersion dominance will be explored in Sections 4.2 (on bi-polarization) and 4.3 (on tail-heaviness).

Recognition that the dispersion curves, $D(t)$, are dual measures suggests that, in seeking a summary measure of dispersion, application of the dual independence axiom (defined on the set of dispersion curves) will lead to measures of the form:

$$(4.3) \quad \Delta_c = \int_0^1 c(t)D(t)dt ,$$

where $c(t)$ is a positive weighting function and linearity implies that it depends only on t . Thus, Δ_c measures the extent of dispersion in a distribution function F .

Where $c(t)=1$ for all t

$$(4.4) \quad \Delta_c = \Delta \equiv \frac{E(X|X \geq M) - E(X|X \leq M)}{M} .$$

By considering the sum of the head count measures defined by (2.6) and (3.5) we find that $\Delta_I = \Pi_I + \Gamma_I$, which means that the dispersion measure Δ_I is equal to the sum of the poverty and affluence head counts. Since Δ_I preserves neither second-degree upwards dispersion dominance nor second-degree downwards dispersion dominance, it appears attractive to use the criteria of second-degree upward and downward dispersion dominance as a basis for imposing further restrictions on the weighting function c . As will be demonstrated in Sections 4.2 and 4.3 the distinction between upward and downward dispersion dominance explains why polarization and tail-heaviness can be considered as complementary concepts of dispersion. As an alternative to the Δ_c family of dispersion measures we get, by relying on the independence axiom rather than the dual independence axiom, the following family of primal measures of dispersion,

$$(4.5) \quad \Theta_e = \int_0^\infty e(z)N(z)dz ,$$

where $e(z)$ is a positive weighting function.

4.2. Bi-Polarization

As we noted at the outset, the distinction between transfers across the median and those on one side of the median was drawn clearly by Foster and Wolfson (1992/2010). It is on this basis that they distinguish between inequality and polarization:

“the distinction between inequality and polarization, earlier noted by Love and Wolfson (1976), may be more clearly identified. ... polarization and inequality move in the same direction when the transfer takes place across the middle ... However, increased bipolarity is associated with a pair of *progressive* transfers, one on each side of the middle, which

necessarily *diminish* inequality. Polarization and inequality move in opposite directions when same-side transfers occur” (2010, pages 251-252, their italics).

In terms of the graphical representation, we can see from Figure 1 that an across-median transfer brings both the $H(z)$ and $K(z)$ curves closer to the vertical line at 1. If everyone had the median income, then the curves would simply coincide with that line. The effect of transfers on one side of the median is less clear. As may be seen from Figure 2, such transfers bring the $H(z)$ and $K(z)$ curves closer at the bottom but further away at the top. Is it evident that this increases polarization? As shown, the transfers certainly build up mass in the frequency distribution around the crossing-point, and this is the basis for the bi-polarity argument of Foster and Wolfson (1992/2010, Figure 2). However, the massing point depends on the location of the transfers. Transfers that are all across a specified value of z will increase mass at that point and, in that sense, increase polarization, but other below-median transfers may move mass away from that point. An argument that does not have this property is that based by Foster and Wolfson on the integration of the $H(z)$ and $K(z)$ curves. In their definition of the second-degree dominance condition, they integrate in each case from $z = 1$, downwards in the case of $H(z)$ and upwards in the case of $K(z)$. This is the reverse of the direction of integration applied for the measures of poverty and affluence in Sections 2 and 3, and explains why the ranking is reversed. In terms of the right hand side of Figure 2, Foster and Wolfson are saying that area A is more important than area B.

As is clear from the above discussion, bi-polarization might be assumed to increase as a result of an application of a combination of below and above median progressive transfers (see Definitions 2.3 and 3.2) and mean-median-preserving across-median regressive transfers (see Definition 2.1). These transfer principles are analogous to the axioms of Increased Bipolarity and Increased Spread introduced by Wang and Tsui (2000) and the axioms of Within-Group Clustering and Between-Group Clustering introduced by Bossert and Schworm (2008). These two pairs of axioms were in both cases used to characterize polarization. The following result demonstrates that second-degree upward dispersion dominance is associated with bi-polarization,

Theorem 4.1. Let D_1 and D_2 be members of the family \mathbf{D} of dispersion curves. Then the following statements are equivalent,

- (i) D_1 second-degree upward dominates D_2 .
- (ii) D_1 can be attained from D_2 by application of a combination of below and above median regressive transfers and/or mean-median-preserving across-median progressive transfers.

The proof of Theorem 4.1 is omitted since it is analogue to the proof of Theorem 2.1. Theorem 4.1 justifies the function P defined by

$$(4.6) \quad P(u) = \int_0^u D(t)dt, \quad u \in [0,1],$$

as a device for comparing bi-polarization between distribution functions. Accordingly, P is denoted *the bi-polarization curve*. The following alternative expression for P

$$(4.7) \quad P(u) = \frac{u}{M} \left[E \left(X \mid M \leq X \leq F^{-1} \left(\frac{1+u}{2} \right) \right) - E \left(X \mid F^{-1} \left(\frac{1-u}{2} \right) \leq X \leq M \right) \right], \quad u \in [0,1],$$

provides an intuitive justification for why it makes sense to consider P as a bi-polarization curve. Assume that F_1 and F_2 are income distributions with corresponding bi-polarization curves P_1 and P_2 , where $P_1(u) \leq P_2(u)$ for all $u \in [0,1]$ and the inequality is strict for at least one $u \in (0,1)$. Then we say

that F_1 exhibits less bi-polarization than F_2 . Note that bi-polarization dominance is equivalent to second-degree upward dispersion dominance. Moreover, by introducing the normalization condition $c(1) = 0$ for the dispersion measures Δ_c defined by (4.3), we get by using integration by parts and inserting (4.4) that Δ_c can be given the following alternative expression in terms of the bi-polarization curve P ,

$$(4.8) \quad \Delta_c = -\int_0^1 c'(u)P(u)du,$$

The use of Δ_c as a measure of bi-polarization requires that $c(t)$ is decreasing, which is justified by the following result,

Theorem 4.2. Let D_1 and D_2 be members of \mathcal{D} . Then the following statements are equivalent,

- (i) D_1 second-degree upward dominates D_2 .
- (ii) $\Delta_c(D_1) < \Delta_c(D_2)$ for all non-negative c such that $c'(t) < 0$ for all $t \in \langle 0, 1 \rangle$.

(Proof in Appendix 2).

To ensure equivalence between bi-polarization dominance (second-degree upward dispersion dominance) and Δ_c -measures as ranking criteria, Theorem 4.2 shows that it is necessary to restrict the weighting functions c to be decreasing, which means that increases of spread in the central part of the distribution receive higher weight than increases of spread at the tails. This feature becomes clear for Δ_2 when $c(t) = 2(1-t)$ since

$$(4.9) \quad \begin{aligned} \Delta_c &= \Delta_2^* \equiv \Delta_1 - \frac{\mu_l}{M} G_l - \frac{\mu_u}{M} G_u = 2\Delta_1 - 4 \frac{\mu}{M} G \\ &= \frac{1}{M} \left[E \left(\min_{i=1,2} X_i \mid X_i \geq M, i=1,2 \right) - E \left(\max_{i=1,2} X_i \mid X_i \leq M, i=1,2 \right) \right], \end{aligned}$$

where μ_l and G_l are the mean and the Gini coefficient of the distribution of incomes below the median and μ_u and G_u are the mean and the Gini coefficient of the distribution of incomes above the median, $G = \frac{\mu_l}{4\mu} G_l + \frac{\mu_u}{4\mu} G_u + \frac{M}{\mu} \Delta_1$ is the Gini coefficient of the income distribution F and

X_1 and X_2 are independent random variables with distribution F . Note that Foster and Wolfson (2010) have provided an alternative justification for Δ_2^* . Inserting for Δ_1 , Π_2 and Γ_2 in (4.8) yields $\Delta_2^* = 2\Delta_1 - (\Pi_2 + \Gamma_2)$, which demonstrates that decreasing poverty and affluence leads to a rise in the bi-polarization measure Δ_2^* provided that the gap between the upper and lower means is kept fixed.

By introducing the family of weighting functions defined by

$$(4.10) \quad c_k(t) = k(1-t)^{k-1}, \quad k = 2, 3, \dots$$

we obtain the following family of bi-polarization measures

$$\begin{aligned}
\Delta_k^* &= \frac{k2^k}{M} \int_{\frac{1}{2}}^1 (1-t)^{k-1} F^{-1}(t) dt - \frac{k2^k}{M} \int_0^{\frac{1}{2}} t^{k-1} F^{-1}(t) dt \\
(4.11) \quad &= \frac{1}{M} \left[E\left(\min_{i \leq k} X_i \mid X_i \geq M, i = 1, 2, \dots, k\right) - E\left(\max_{i \leq k} X_i \mid X_i \leq M, i = 1, 2, \dots, k\right) \right], k = 2, 3, \dots
\end{aligned}$$

Note that $\Delta_k^* \rightarrow 0$ when $k \rightarrow \infty$, which means that Δ_k^* approaches bi-polarization neutrality when $k \rightarrow \infty$.

An interesting question is whether Δ_k^* for $k > 2$ is related to summary measures of inequality in a similar way as Δ_2^* is related to the lower and upper tail Gini coefficients. By noting that the upper tail version of the extended Gini family of inequality measures (Donaldson and Weymark, 1980) and the lower tail version of the Lorenz family of inequality measures (Aaberge, 2000) are given by

$$(4.12) \quad G_{u,k} = 1 - \frac{k2^k}{\mu_u} \int_{\frac{1}{2}}^1 (1-t)^{k-1} F^{-1}(t) dt, \quad k \geq 2$$

and

$$(4.13) \quad D_{l,k} = \frac{k2^k}{(k-1)\mu_l} \int_0^{\frac{1}{2}} t^{k-1} F^{-1}(t) dt - \frac{1}{k-1}, \quad k \geq 2,$$

we get by inserting (4.12) and (4.13) into (4.10)

$$(4.14) \quad \Delta_k^* = \Delta - (k-1) \frac{\mu_l}{M} D_{l,k} - \frac{\mu_u}{M} G_{u,k}, \quad k \geq 2,$$

where $G_{l,2}$ and $G_{u,2}$ represents the lower and upper tail Gini coefficients. Since the extended Gini family is designed to focus attention on changes that place in lower part, and the Lorenz family on the upper part of the income distribution, the upper tail extended Gini coefficients will be particularly sensitive to transfers that affect income just above the median and similarly the members of the lower tail Lorenz family will be particularly sensitive to transfers that concern incomes that are slightly smaller than the median.

The measures of bi-polarization described so far have been based on the dual approach: considering the income gap between people located at the same percentile on each side of the median. The primal approach considers the gap between the proportions of the population having incomes a certain fraction away from the median income. The integral of this distance, shown as $N(z)$ in Figure 6, can provide the basis for alternative classes of bi-polarisation measures, with weights that are decreasing functions of z .

As an alternative approach for measuring polarization, Esteban and Ray (1994, 1999) and Duclos et al. (2004) adopt an identification-alienation framework and start from a general family of polarization measures defined as a functional of the probability density. By requiring that the general family of polarization measures satisfies four axioms, Duclos et al. (2004) obtained a characterization of a parametric sub-family of polarization measures. As was recognized by Duclos et al (2004), several other measures of polarization might satisfy their four axioms. Actually, it can be demonstrated that the criterion of bi-polarization dominance proposed in this paper satisfies these axioms. However, we also find that the criterion of (first-degree) dispersion dominance satisfies

Axioms 1, 3 and 4, whereas Axiom 2 is only satisfied by second-degree upward dispersion dominance which is equivalent to first-degree bi-polarization dominance. Duclos et al. acknowledge the importance of Axiom 2 by stating “In some sense, this is the defining axiom of polarization, and may be used to motivate the concept” (2004, page 1742). The three remaining axioms are useful conditions for measuring dispersion, but do not discriminate between bi-polarization and tail-heaviness. Note that the family of bi-polarization measures Δ_c defined by (4.7) is completely characterized by a bi-polarization curve ordering that is continuous, transitive and complete and satisfies the dual independence axiom and first-degree bi-polarization dominance⁸.

4.3. Tail-heaviness

In constructing the bi-polarisation curve, we integrated outwards from the median. We now consider the implications of integrating from the tails inwards. This links naturally to the concept of tail-heaviness. The definition of tail-heaviness given, for instance, by Doksum (1969, page 1169) involves the limiting behaviour as t goes to 1 of the relative income levels corresponding to a specified quantile. This suggests that tail-heaviness might be associated with the criterion of second-degree downward dispersion dominance and the corresponding theorem to Theorem 4.1. This shows that above and below median progressive transfers in combination with mean-median-preserving across-median progressive transfers provide a characterization of second-degree downward dispersion dominance,

Theorem 4.3. Let D_1 and D_2 be members of the family \mathbf{D} of dispersion curves. Then the following statements are equivalent,

- (i) D_1 second-degree downward dominates D_2
- (ii) D_1 can be attained from D_2 by application of a combination of below and above median progressive transfers and mean-median-preserving across-median progressive transfers.

Note that both tail-heaviness and bi-polarization increase as a result of mean-median-preserving across-median progressive transfers, whereas tail-heaviness decreases and bi-polarization increases as a consequence of application of below and above median progressive transfers. In this sense, tail-heaviness fits more naturally with measures of inequality.

The equivalent result to Theorem 4.2 provides a justification for using the function T defined by

$$(4.15) \quad T(u) = \int_u^1 D(t)dt, \quad u \in [0,1],$$

as a basis for comparing distribution functions with respect to tail-heaviness. Accordingly, we denote T the *tail-heaviness curve*. The following alternative expression for T

$$(4.16) \quad T(u) = \frac{(1-u)}{M} E\left(X \middle| X \geq F^{-1}\left(\frac{1+u}{2}\right)\right) - E\left(X \middle| X \leq F^{-1}\left(\frac{1-u}{2}\right)\right), \quad u \in [0,1],$$

⁸ The papers of the special issue “Income Polarization: Measurement, Determinants, and Implications” of *Review of Income and Wealth* edited by Deutsch and Silber (2010) discuss theoretical as well as empirical issues concerning the concept of polarization. See also Chakravarty (2009).

gives an intuitive explanation for why T can be considered as a tail-heaviness curve. Assume that F_1 and F_2 are income distributions with corresponding tail-heaviness curves T_1 and T_2 , where $T_1(u) \leq T_2(u)$ for all $u \in [0,1]$ and the inequality is strict for at least one $u \in (0,1)$. Then we say that F_1 exhibits less tail-heaviness than F_2 .

Expression (4.15) shows that tail-heaviness dominance is equivalent to second-degree downward dispersion dominance. By introducing the weighting function $c(u)$, with the normalization condition $c(1) = 0$, it can be demonstrated that the family of dispersion measures Δ_c can be expressed in terms of the tail-heaviness curve T ,

$$(4.17) \quad \Delta_c = \int_0^1 c'(u) T(u) du.$$

To employ Δ_c as a summary measure of tail-heaviness expression (4.17) suggests that $c(u)$ should be an increasing function of u , which is confirmed by the following result,

Theorem 4.4. Let D_1 and D_2 be members of \mathbf{D} . Then the following statements are equivalent,

- (i) D_1 second-degree downward dominates D_2
- (ii) $\Delta_c(D_1) < \Delta_c(D_2)$ for all non-negative c such that $c'(u) > 0$ for all $u \in [0,1]$.

(Proof in Appendix 2).

Let $c(t) = 2t$. Then

$$(4.18) \quad \Delta_c = \Delta_2^{**} \equiv \Delta_l + \frac{\mu_l}{M} G_l + \frac{\mu_u}{M} G_u = \frac{4\mu}{M} G,$$

where G_l is G_u the Gini coefficients of the conditional income distributions given that incomes takes values below and above the median, respectively. Similarly as for Δ_l we find that Δ_2^{**} can be expressed as the sum of the associated poverty and affluence measures; i.e. $\Delta_2^{**} = \Pi_2 + \Gamma_2$. By introducing the family of weighting functions defined by

$$(4.19) \quad c_k(t) = kt^{k-1}, k = 2, 3, \dots,$$

we obtain the following family of tail-heaviness measures

$$(4.20) \quad \begin{aligned} \Delta_k^{**} &= \frac{2k}{M} \int_{\frac{1}{2}}^1 (2t-1)^{k-1} F^{-1}(t) dt - \frac{2k}{M} \int_0^{\frac{1}{2}} (1-2t)^{k-1} F^{-1}(t) dt \\ &= \frac{1}{M} \left[E \left(\max_{i \leq k} X_i \mid X_i \geq M, i = 1, 2, \dots, k \right) - E \left(\min_{i \leq k} X_i \mid X_i \leq M, i = 1, 2, \dots, k \right) \right], k = 2, 3, \dots \end{aligned}$$

Note that

$$(4.21) \quad \Delta_k^{**} \rightarrow \frac{F^{-1}(1) - F^{-1}(0+)}{M} \text{ when } k \rightarrow \infty.$$

As for Δ_2^{**} we find that Δ_k^{**} for $k > 2$ can be expressed as the sum of associated measures of poverty and affluence, combining the results of Sections 2 and 3:

$$(4.22) \quad \Delta_k^{**} = \Pi_k + \Gamma_k, k \geq 2.$$

Inserting (2.11) and (3.11) for Π_k and Γ_k in (4.22) yields

$$(4.23) \quad \Delta_k^{**} = \Delta_1 + \frac{\mu_l}{M} G_{l,k} + \frac{\mu_u}{M} D_{u,k}, k \geq 2.$$

As opposed to the bi-polarization measures Δ_k^* for $k > 2$, the tail-heaviness measures Δ_k^{**} for $k > 2$ are more sensitive to changes that take place at the tails than at the central part of the distribution function. Moreover, the sensitivity to changes at the tails increases with increasing k .

A specific family of primal tail-heaviness measures is obtained by using the weighting function $e(z) = kz^{k-1}$ in (4.5),

$$(4.24) \quad \Theta_e \equiv \Theta_k = \int_0^\infty e(z) N(z) dz = \frac{E[(M - X)^k | X \leq M] + E[(X - M)^k | X \geq M]}{M^k}, k \geq 2.$$

Moreover, by inserting from (2.12) and (3.12) into (4.24) we find that Θ_e can be expressed as the sum of associated measures of poverty and affluence,

$$(4.24) \quad \Theta_k = \Psi_k + \Lambda_k, k \geq 2.$$

In this way, we have come full circle.

5. Conclusions

In this paper, we have considered the measurement of poverty, affluence, bi-polarization and tail-heaviness in a society that is both well-off and has a middle class, which we interpret to mean that the median person is neither poor nor rich. In such a society, the median operates like a “watershed”:

- Transfers from people above the median to people below the median reduce inequality, reduce poverty, reduce the proportion classed as “rich”, and reduce bi-polarization and tail-heaviness;
- Progressive transfers within the bottom half of the population reduce inequality; they may either reduce or raise the poverty headcount, but definitely reduce the poverty gap and other concave measures of poverty; these transfers have no impact on the proportion classed as “rich”; they reduce tail-heaviness but increase bi-polarization;

- Progressive transfers within the top half of the population reduce inequality, and have no impact on poverty; these transfers may either reduce or raise the proportion classed as “rich”, but definitely reduce the affluence gap; these transfers reduce tail-heaviness but increase bi-polarization.

These conclusions are summarized in Table 1 (where F/W refers to Foster and Wolfson).

Table 1 Effect of progressive mean- median-preserving transfers

	1.	2.	3.	4.	5.
	Inequality	Poverty	Affluence	Bi-Polarization	Tail-heaviness
Across median	Fall	Fall	Fall	Fall	Fall
Below median	Fall	Fall where weighting function non-increasing	Not relevant	Rise (F/W)	Fall
Above median	Fall	Not relevant	Fall where weighting function non-decreasing	Rise (F/W)	Fall

The aim of the paper has been to provide a unified framework for examining different distributional measures, bringing out the ways in which they are different and the extent to which they have a common structure. In particular,

- By building systematically from dominance results towards specific poverty measures it is possible to overcome some of the criticisms made of standard measures.
- It is possible to define measures of affluence parallel to those of poverty with similar properties, but with appropriate transpositions (such as integrating downwards, giving more weight to transfers at the top).
- Bi-Polarisation and tail-heaviness are both measures of dispersion, but with the crucial difference that one starts from the median and gives more weight to transfers near the middle, whereas the other starts from the extremes and gives more weight to transfers far removed from the median.
- The concept of tail-heaviness fits more naturally with measures of inequality.
- Certain elements are common to all aspects, notably the weighting functions and the implications of different assumptions for the relation with independence axioms.
- The common framework helps understand the relation between different measures used in the literature, such as different measures of poverty and different approaches to measuring bi-polarisation.

We have argued that the conclusions drawn with respect to all five concepts - poverty, affluence, inequality, bi-polarization and tail-heaviness – depend on the perspective taken when

examining the distribution of income (or other variables such as consumption). All five concepts can be viewed from either a primal (looking at incomes) or a dual (looking at ranks) perspective; this affects both dominance conditions and specific measures. Well-known measures can be seen as linked by a duality relationship: for example, in the case of poverty indicators the Sen measure is a dual measure and the Foster-Greer-Thorbecke measure is of the primal form.

Appendix 1: Asymptotic estimation theory

This appendix deals with the asymptotic distributions of the non-parametric estimators of the summary measures Ψ_a , Π_p , A_b , Γ_q , Δ_c and Θ_e .

Let X be an income variable with cumulative distribution function F and mean μ . Let $[a, b]$ be the domain of F where F^{-1} is the left inverse of F and $F^{-1}(0) \equiv a \geq 0$.

Let X_1, X_2, \dots, X_n be independent random variables with common distribution function F and let F_n be the corresponding empirical distribution function and let $M_n = F_n^{-1}(0.5)$. Since the parametric form of F is not known, it is natural to use the empirical distribution function F_n to estimate F , to use $H_n(z) = 2F_n(M_n z)$ to estimate $H(z)$, and to use $K_n(z) = 2(1 - F_n(M_n z))$ to estimate $K(z)$. Since F_n and M_n are consistent estimators of F and M , $H_n(z)$ and $K_n(z)$ are consistent estimators of $H(z)$ and $K(z)$.

Let the empirical processes $P_n(x)$ and $Q_n(t)$ be defined by

$$P_n(x) = n^{\frac{1}{2}} (F_n(x) - F(x)) \text{ and } Q_n(t) = n^{\frac{1}{2}} (F_n^{-1}(t) - F^{-1}(t))$$

and let $W_0(t)$ denote a Brownian Bridge on $[0, 1]$, that is, a Gaussian process with mean zero and covariance function $s(1-t)$, $0 \leq s \leq t \leq 1$.

Since the support of F is a non-empty finite interval $[a, b]$ (when F is an income distribution, a is commonly equal to zero), $P_n(x)$ is member of the space A of functions on $[a, b]$, and $Q_n(t)$ is member of the space B of functions on $[0, 1]$ which are right continuous and have left hand limits. On these spaces we use the Skorokhod topology and the associated σ -field (e.g. Billingsley 1968, page 111).

Let $v(x)$ be a positive non-increasing or non-decreasing function of x and let $0 \leq a < b < \infty$. In order to study the asymptotic behaviour of the empirical counterparts of Ψ_a , A_b and Θ_e it is convenient to consider the empirical process

$$V_n = \int_a^b v\left(\frac{x}{M_n}\right) P_n(x) dx.$$

Proposition A.1a. Suppose that F has a continuous nonzero derivative f on $[a, b]$. Then $V_n(u)$ converges in distribution to the process

$$V = \int_a^b v\left(\frac{x}{M}\right) W_0(F(x)) dt .$$

Proof. By a well-known result (see e.g. Billingsley, 1968, p 141) $P_n(x)$ converges in distribution to $W_0(F(x))$. Using the arguments of Durbin (1973, Section 4.4), we find that V as a function of $W_0(F(x))$ is continuous in the Skorokhod topology. Then it follows from Billingsley (1968, Theorem 5) that

$$\int_a^b v\left(\frac{x}{M}\right) P_n(x) dt \text{ converges in distribution to } V.$$

Since M_n converges in probability to M , Cramer-Slutsky's theorem gives that $V_n(u)$ converges in distribution to V . \square

Let $w(t)$ be a positive non-increasing or non-decreasing function of x and let $0 \leq m < r \leq I$. In order to study the asymptotic behaviour of the empirical counterparts of Π_p , Γ_q and Δ_c it is convenient to consider the empirical process

$$Y_n = \int_m^r w(t) Q_n(t) dt ,$$

Proposition A.1b. Suppose that F has a continuous nonzero derivative f on $[a, b]$. Then $Y_n(u)$ converges in distribution to the process

$$Y = \int_m^r w(t) \frac{W_0(t)}{f(F^{-1}(t))} dt .$$

Proof. It follows from Theorem 4.1 of Doksum (1974) that the empirical process $Q_n(t)$ converges in distribution to the Gaussian process $W_0(t)/f(F^{-1}(t))$. Using the arguments of Durbin (1973, Section 4.4), we find that Y as function of $W_0(t)/f(F^{-1}(t))$ is continuous in the Skorokhod topology. The results then follow from Billingsley (1968, Theorem 5.1). \square

Let $\Delta_{c,n}$ be the empirical counterpart of Δ_c . By using arguments like those in the proof of Proposition A.1b we get the following result,

Proposition A.2. Suppose that F has a continuous nonzero derivative f on $[a, b]$. Then $\Delta_{c,n}$ converges in distribution to the process

$$\frac{1}{M} b \int_M^b c(2F(x) - 1) W_0(F(x)) dx - \frac{1}{M} \int_a^M c(1 - 2F(x)) W_0(F(x)) dx.$$

Propositions A.1a, A.1b and A.2 demonstrate that V , Y and Δ_c are functionals of Gaussian processes and thus that V_n , Y_n and $\Delta_{c,n}$ are asymptotically normally distributed.

Appendix 2: Proofs of theoretical results

LEMMA 1. Let M be the family of bounded, continuous and non-negative functions on $[0, 1]$ which are positive on $\langle 0, 1 \rangle$ and let g be an arbitrary bounded and continuous function on $[0, 1]$. Then

$$\int g(t) h(t) dt > 0 \text{ for all } h \in M$$

implies

$$g(t) \geq 0 \text{ for all } t \in [0, 1]$$

and the inequality holds strictly for at least one $t \in \langle 0, 1 \rangle$.

The proof of Lemma 1 is known from mathematical textbooks.

Proof of Theorem 2.1. We start with proving the equivalence between (i) and (iii). Using integration by parts we have that

$$0 < \Psi_a(H_2) - \Psi_a(H_1) = a(1) \int_0^1 (H_2(z) - H_1(z)) dz - \int_0^1 a'(y) \int_0^y (H_2(z) - H_1(z)) dz dy$$

Thus, if (i) holds then $\Psi_a(H_1) < \Psi_a(H_2)$ for all decreasing a such that $a(1) = 0$.

To prove the converse statement we restrict to decreasing a such that $a(1) = 0$. Hence,

$$0 < \Psi_a(H_2) - \Psi_a(H_1) = - \int_0^1 a'(y) \int_0^y (H_2(z) - H_1(z)) dz dy$$

and the desired result is obtained by applying Lemma 1.

Next, to prove the equivalence between (i) and (iv) note that

$$\int_0^y (H_2(z) - H_1(z)) dz \geq 0 \text{ for all } y \in [0, I]$$

is equivalent to the following condition (see Atkinson, 1970)

$$\int_0^u (H_1^{-1}(t) - H_2^{-1}(t)) dt \geq 0 \text{ for all } u \in [0, I].$$

Then, using integration by parts we get that

$$0 < \Pi_p(H_2) - \Pi_p(H_1) = p(I) \int_0^I (H_1^{-1}(t) - H_2^{-1}(t)) dt - \int_0^I p'(u) \int_0^u (H_1^{-1}(t) - H_2^{-1}(t)) dt du.$$

Thus, if (i) holds then $\Pi_a(H_1) < \Pi_a(H_2)$ for all decreasing p such that $p(1) = 0$.

To prove the converse statement we restrict to decreasing p such that $p(1) = 0$. Hence,

$$0 < \Pi_p(H_2) - \Pi_p(H_1) = - \int_0^I p'(u) \int_0^u (H_1^{-1}(t) - H_2^{-1}(t)) dt du,$$

and the desired result is obtained by applying Lemma 1.

To prove the equivalence between (ii) and (iii) consider a case where we transfer a small amount γ from a person with income $F^{-1}(t+h)$ to a person with income $F^{-1}(t)$, where $0 < t < t+h < 0.5$. Then Ψ_a defined by (2.1) decreases if and only if

$$a(t) - a(t+h) > 0$$

which for small h is equivalent to

$$a'(z) < 0.$$

□

The proofs of Theorem 3.1 and 4.2 are analogous to the proof of Theorem 2.1 and are based on the expressions

$$A_b(K_2) - A_b(K_1) = \int_1^\infty b(z) (K_2(z) - K_1(z)) dz = \int_1^\infty b'(y) \int_y^\infty (K_2(z) - K_1(z)) dz dy,$$

$$\Gamma_q(K_2) - \Gamma_q(K_1) = \int_0^1 q(t) (K_2^{-1}(1-t) - K_1^{-1}(1-t)) dt = \int_0^1 q'(1-u) \int_0^u (K_2^{-1}(t) - K_1^{-1}(t)) dt du$$

and

$$A_c(D_2) - A_c(D_1) = \int_0^1 c(t) (D_2(t) - D_1(t)) dt = - \int_0^1 c'(u) \int_0^u (D_2(t) - D_1(t)) dt du,$$

which are obtained by using integration by parts. Thus, by arguments like those in the proof of Theorem 2.1 the results of Theorem 3.1 and 4.2 are obtained.

References

- Aaberge, R. (2000): Characterizations of Lorenz curves and income distributions, *Social Choice and Welfare* **17**, 639-653.
- Aaberge, R. (2001): "Axiomatic characterization of the Gini coefficient and Lorenz curve orderings", *Journal of Economic Theory*, 101, 115-132.
- Aaberge, R. (2009): "Ranking intersecting Lorenz curves", *Social Choice and Welfare*, **33**, 235-259.
- Aaberge, R. and M. Mogstad (2010): "On the measurement of long-term income inequality and income mobility", IZA Discussion Paper 4699.
- Atkinson, A. B. (1970): "On the measurement of inequality", *Journal of Economic Theory*, **2**, 244-263.
- Atkinson, A. B. (1987): "On the measurement of poverty", *Econometrica*, **55**, 749-764.
- Bickel, P. J. and E.H. Lehmann (1979): "Descriptive statistics for nonparametric models". IV. Spread, in J. Jureckova, editor, *Contributions to Statistics*, Reidel: Dordrecht.
- Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York.
- Bossert, W. and W. Schworm (2008): "A class of two-group polarization measures", *Journal of Public Economic Theory*, **10**, 1169-1187.
- Brzezinski, M. (2010): "Income Affluence in Poland", *Social Indicators Research*, **99**, 285-299.
- Chakravarty, S. R. (2009): *Inequality, Polarization and Poverty*, Springer: New York.
- Deutsch, J. and J. Silber (2010): "Income polarization: Measurement, determinants, and implications", *Review of Income and Wealth*, 56, 1-6.
- Doksum, K. (1969): "Starshaped transformations and the power of rank tests", *Annals of Mathematical Statistics*, **40**, 1167 – 1176.
- Donaldson, D. and J.A. Weymark (1980): A single parameter generalization of the Gini indices of inequality, *Journal of Economic Theory* **22**, 67-86.
- Duclos, J.-Y., J. Esteban and D. Ray (2004): "Polarization: Concepts, measurement, estimation", *Econometrica*, **72**, 1737-1772.
- Durbin, J. (1973): *Distribution Theory for Tests Based on the Sample Distribution Function*. Society for Industrial and Applied Mathematics, Philadelphia.
- Esteban, J. and D. Ray (1994): "On the measurement of polarization", *Econometrica*, **62**, 819-851.
- Esteban, J. and D. Ray (1999): "Conflict and distribution", *Journal of Economic Theory*, **87**, 379-415.
- Esteban, J. and D. Ray (2012): "Comparing polarization measures" in M. Garfinkel and S. Skaperdas, editors, *Oxford Handbook of the Economics of Peace and Conflict*, Oxford University Press, Oxford: 127-151.

- Foster, J. E., J. Greer and E. Thorbecke (1984): "A class of decomposable poverty measures", *Econometrica*, **52**, 761-766.
- Foster, J. E. and M. C. Wolfson (1992/2010): "Polarization and the decline of the middle class: Canada and the U.S.", *Journal of Economic Inequality*, **8**, 247-273.
- Green, J. R. and B. Jullien (1988): "Ordinal independence in nonlinear utility theory", *Journal of Risk and Uncertainty*, **1**, 355-387.
- Jenkins, S. P. and P. J. Lambert (1997): "Three 'I's of Poverty Curves, with an Analysis of UK Poverty Trends", *Oxford Economic Papers*, 317-327.
- Peichl, A., T. Schaefer and C. Scheicher (2010): "Measuring Richness and Poverty: A Micro Data Application to Europe and Germany", *Review of Income and Wealth*, **56**, 597-619.
- Sen, A. (1976): "Poverty: An ordinal approach to measurement", *Econometrica*, **44**, 219-231.
- Sen, P. K. (1988): "The harmonic Gini coefficient and affluence indexes", *Mathematical Social Sciences*, **16**, 65-76.
- Shorrocks, A. (1995): "Revisiting the Sen poverty index", *Econometrica*, **63**, 1225-1230.
- Stiglitz, J. E., A. Sen and J.-P., Fitoussi (2009): *Report by the Commission on the Measurement of Economic Performance and Social Progress*,
- Wang, Y.-Q. and K.-Y., Tsui (2000): "Polarization orderings and new classes of polarization indices", *Journal of Public Economic Theory*, **2**, 349-363.
- Wolfson, M. (1994): "When inequalities diverge", *American Economic Review*, **84**, 353-358.
- Yaari, M.E. (1987): "The dual theory of choice under risk", *Econometrica*, **55**, 95-115.
- Yaari, M.E. (1988): "A controversial proposal concerning inequality measurement", *Journal of Economic Theory*, **44**, 381-397.