POLARIZATION: CONCEPTS, MEASUREMENT, ESTIMATION[1]

By Jean-Yves Duclos, Joan Esteban and Debraj Ray[2]

We develop the measurement theory of polarization for the case in which income distributions can be described using density functions. The main theorem uniquely characterizes a class of polarization measures that fit into what we call the "identity-alienation" framework, and simultaneously satisfies a set of axioms. Second, we provide sample estimators of population polarization indices that can be used to compare polarization across time or entities. Distribution-free statistical inference results are also used in order to ensure that the orderings of polarization across entities are not simply due to sampling noise. An illustration of the use of these tools using data from 21 countries shows that polarization and inequality orderings can often differ in practice.

Keywords: polarization, inequality, conflict, group identity, axiomatics

1. introduction

Initiated by Esteban and Ray (1991, 1994), Foster and Wolfson (1992) and Wolfson (1994), there has been a recent upsurge of interest in the measurement of polarization[3] and in the use of such measures as a correlate of different aspects of socioeconomic performance. It seems fairly widely accepted that polarization is a concept that is distinct from inequality,

[1]In this chapter, we reproduce our paper published in Econometrica 2004, with an addendum containing the proofs of the estimation and statistical inference results that are reported in section 3.1.

[2]This research was funded by The Pew Charitable Trusts, CRSH, FQRSC, the Chair of Canada in Social Policies and Human Resources, and National Science Foundation 0241070 (Ray). Duclos and Ray thank the Instituto de Análisis Económico (CSIC) for hospitality during the startup phase of this project. Esteban is a member of Barcelona Economics and thanks the support from the Generalitat de Catalunya, the Instituto de Estudios Fiscales and the MCYT. We thank Oliver Linton, Patrick Richard, a Co-Editor, and two anonymous referees for useful comments. Finally, we are grateful to Nicolas Beaulieu for his excellent research assistance.

[3]See Esteban and Ray (1991, 1994), Foster and Wolfson (1992), Wolfson (1994, 1997), Alesina and Spolaore (1997), Quah (1997), Wang and Tsui (2000), Esteban, Gradín and Ray (1998), Chakravarty and Majumder (2001), Zhang and Kanbur (2001) and Rodríguez and Salas (2002).

and that — at least in principle — it could be connected with several aspects of social, economic and political change.[4]

Following Esteban and Ray (1991, 1994), we rely almost exclusively on what might be called the *identification-alienation* framework. The idea is simple: polarization is related to the alienation that individuals and groups feel from one another, *but such alienation is fuelled by notions of within-group identity.* In concentrating on such phenomena, we do not mean to suggest that instances in which a single isolated individual runs amok with a machine gun are rare, or that they are unimportant in the larger scheme of things. It is just that these are not the objects of our enquiry. We are interested in the correlates of organized, large-scale social unrest — strikes, demonstrations, processions, widespread violence, and revolt or rebellion. Such phenomena thrive on differences, to be sure. But they cannot exist without notions of group identity either.

This brief discussion immediately suggests that inequality, inasmuch as it concerns itself with interpersonal alienation, captures but one aspect of polarization. To be sure, there are some obvious changes that would be branded as both inequality- and polarization-enhancing. For instance, if two income groups are further separated by increasing economic distance, inequality and polarization would presumably both increase. However, *local* equalizations of income differences at two different ranges of the income distribution will most likely lead to two better-defined groups — each with a clearer sense of itself and the other. In this case, inequality will have come down but polarization may be on the rise.

The purpose of this paper is two-fold. First, we develop the measurement theory of polarization for the case in which the relevant distributions can be described by density functions. There are many such instances, the most important being income, consumption and wealth — regrouped under "income" for short. The reason for doing so is simple: with sample data aggregated along income intervals, it is unclear how to provide a statistically satisfactory account of whether distributive measures (based on such data) are

[4]See, for instance, D'Ambrosio and Wolff (2001), Collier and Hoeffler (2001), Fajnzylber, Lederman and Loayza (2000), Garcia-Montalvo and Reynal-Querol (2002), Gradín (2000), Knack and Keefer (2001), Milanovic (2000), Quah (1997) and Reynal-Querol (2002). See also Esteban and Ray (1999) for a formal analysis of the connections between polarization and the equilibrium level of conflict in a model of strategic interaction.

significantly different across time or entities. Indeed, a rapidly burgeoning literature on the statistics of inequality and poverty measurement shows how to construct appropriate statistical tests for such measures using disaggregated data (see, e.g., Beach and Davidson, (1983), Beach and Richmond (1985), Bishop et al. (1989), Kakwani (1993), Anderson (1996), and Davidson and Duclos (1997, 2000)). A rigorous axiomatic development of the polarization concept in the "density case" is then a prerequisite for proper statistical examination of polarization.

In this paper we concentrate on the axiomatics and estimation of "pure income polarization", that is, of indices of polarization for which individuals identify themselves only with those with similar income levels. [However, Section 4 does contain several preliminary remarks on the broader concept of "social polarization."] With this settled, we turn to issues of estimation. The main problem is how to estimate the size of the groups to which individuals belong. Again, using arbitrary income intervals would appear somewhat unsatisfactory. Instead, we estimate group size non-parametrically using kernel density procedures. A natural estimator of the polarization indices is then given by substituting the distribution function by the empirical distribution function. Assuming that we are using a random sample of independently and identically distributed observations of income, the resulting estimator has a limiting normal distribution with parameters that can be estimated free of assumptions on the true (but unknown) distribution of incomes. Distribution-free statistical inference can then be applied to ensure that the orderings of polarization across entities are not simply due to sampling noise.

It is useful to locate this paper in the context of the earlier step in the measurement of polarization in Esteban and Ray (1994) — ER from now on. The measure derived in ER was based on a discrete, finite set of income groupings located in a continuous ambient space of possible income values. This generated two major problems, one conceptual and the other practical. At the conceptual level we have the drawback that the measure presents an unpleasant discontinuity. This is precisely due to the fact that ER is based on a population distributed over a discrete and distinct number of points.[5] The practical difficulty is that the population is assumed to have *already* been bunched in the relevant

[5]ER (Section 4, p. 846) mention this problem.

groups. This feature rendered the measure of little use for many interesting problems.[6] As mentioned above, the present paper addresses both problems and provides what we hope is a useable measure.

In addition, the main axioms that we use to characterize income polarization are substantially different from ER (though they are similar in spirit). In large part, this is due to the fact that we are dealing with a completely different domain (spaces of densities). We therefore find it of interest that these new axioms end up characterizing a measure of polarization that turns out to be the natural extension of ER to the case of continuous distributions. At a deeper level, there are, however, important differences, such as the different bounds on the "polarization-sensitivity" parameter α that are obtained.

In Section 2 we axiomatically characterize a measure of pure income polarization and examine its properties; this is the conceptual heart of the paper. We then turn in Section 3 to estimation and inference issues for polarization measures and subsequently illustrate the axiomatic and statistical results using data drawn from the Luxembourg Income Study (LIS) data sets for 21 countries. We compute the Gini coefficient and the polarization measure for these countries for years in Wave 3 (1989–1992) and Wave 4 (1994–1997), and find *inter alia* that the two indices furnish distinct information on the shape of the distributions. Section 4 summarizes the results and discusses an important extension. All proofs are relegated to an appendix.

2. MEASURING INCOME POLARIZATION

The purpose of this section is to proceed towards a full axiomatization of income polarization.

2.1. *Starting Point*

The domain under consideration is the class of all continuous (unnormalized) densities in $\mathbb{R}_+$, with their integrals corresponding to various population sizes. Let f be such a density; we are interested in its *polarization* $P(f)$. We first describe the notions of

[6] In Esteban, Gradín and Ray (1998) we presented a statistically reasonable way to bunch the population in groups and thus make the ER measure operational. Yet, the number of groups had to be taken as exogenous and the procedure altogether had no clear efficiency properties.

"alienation" and "identification" for each individual with income located in the support of f.

We presume that an individual located at x feels alienation *vis-a-vis* another located at y, and that this alienation is monotonic in distance $|x-y|$. This notion is commonplace in the literature on the conceptual foundations of inequality (see, *e.g.*, Sen (1997)).

At the same time, the "identity-alienation framework" we adopt (referred to as IA henceforth) emphasizes that alienation *per se* is not the end of the story: for alienation to be translated into effective voice, action, or protest, the individual must — to greater or lesser degree — *identify* with others in society. In this paper, we presume that an individual located at income x experiences a sense of identification that depends on the density at x, $f(x)$.

Taken in a broader context, the identification assumption is obviously quite specific. For instance, one might consider the possibility that individuals have a nondegenerate "window of identification" (though the foundations for the width of such an identification window appear unclear). We address this issue (and others) in our discussion of identification in Section 2.4, but recognize that a full analysis of the behavioral foundations of identification is beyond the scope of this paper.

As in ER, we are interested in the *effective antagonism* of x towards y (under f). In its most abstract form, we may depict this as some nonnegative function

$$T(i, a),$$

where $i = f(x)$ and $a = |x - y|$. It is assumed that T is increasing in its second argument and that $T(i, 0) = T(0, a) = 0$, just as in ER. [This last condition asserts that while the consequences of an isolated individual's sense of alienation might be important, this is not the focus of our exercise.] We take polarization to be proportional to the "sum" of all effective antagonisms:

$$(1) \quad P(F) = \int \int T\left(f(x), |x-y|\right) f(x)f(y)dxdy,$$

This class of measures is neither very useful nor operational, though at this stage it it incorporates the structure of the IA assumptions. In particular, much depends on the choice of the functional form T. In what follows, we place axioms on this starting point so as to pin down this functional form.

2.2. *Axioms*

Densities and Basic Operations. Our axioms will largely be based on domains that are unions of one or more very simple densities f that we will call *basic densities.* These are unnormalized (by population), are symmetric and unimodal, and have compact support.[7]

To be sure, f can be *population rescaled* to any population p by simply multiplying f pointwise by p to arrive at a new distribution pf (unnormalized). Likewise, f can undergo a *slide*. A *slide to the right by* x is just a new density g such that $g(y) = f(y-x)$. Likewise for a slide to the left. And f with mean μ' can be *income rescaled* to any new mean μ that we please as follows: $g(x) = (\mu'/\mu)f(x\mu'/\mu)$ for all x.[8] These operations maintain symmetry and unimodality and therefore keep us within the class of basic densities.

If we think of slides and scalings as inducing a partition of the basic densities, each collection of basic densities in the same element of the partition may be associated with a *root*, a basic density with mean 1 and support $[0, 2]$, with population size set to unity. That is, one can transform any basic density to its root by a set of scalings and slides. [This concept will be important both in the axioms as well as in the main proof.] Two distinct roots differ in "shape", a quality that cannot be transformed by the above operations.

Finally, we shall also use the concept of a *squeeze*, defined as follows. Let f be any basic density with mean μ and let λ lie in $(0, 1]$. A λ-*squeeze* of f is a transformation as follows:

$$(2) \qquad f^\lambda(x) \equiv \frac{1}{\lambda} f\left(\frac{x - [1-\lambda]\mu}{\lambda}\right).$$

A (λ-) squeeze is, in words, a very special sort of mean-preserving reduction in the spread of f. It concentrates more weight on the *global* mean of the distribution, as opposed to what would be achieved, say, with a progressive Dalton transfer on the same side of the mean. Thus a squeeze truly collapses a density inwards towards its global mean. The following properties can be formally established: (a) For each $\lambda \in (0, 1)$, f^λ is a density; (b) for each $\lambda \in (0, 1)$, f^λ has the same mean as f; (c) If $0 < \lambda < \lambda' < 1$, then f^λ

[7]By symmetry we mean that $f(m - x) = f(m + x)$ for all $x \in [0, m]$, where m is the mean and by unimodality we mean that f is nondecreasing on $[0, m]$.

[8]The reason for this particular formulation is best seen by examining the corresponding cumulative distribution functions, which must satisfy the property that $G(x) = F(x\mu'/\mu)$, and then taking derivatives.

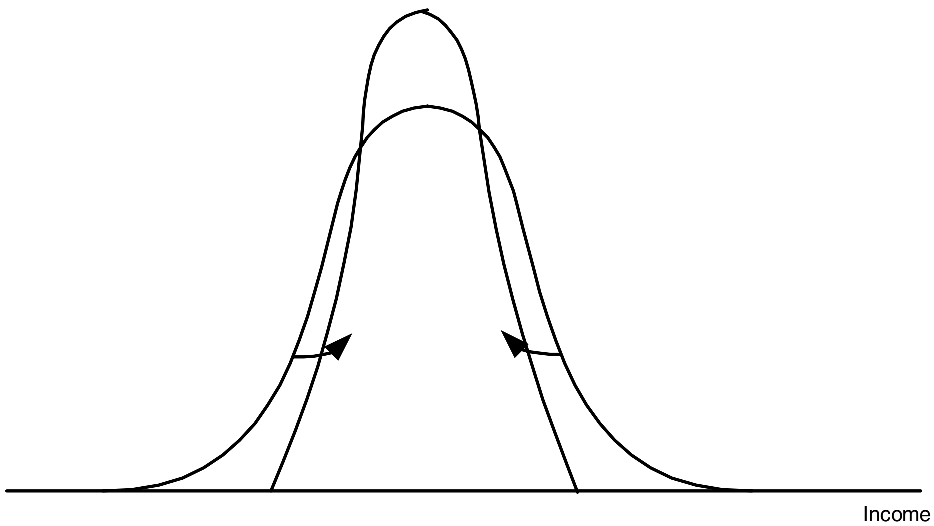

Figure 1: —A Single Squeeze Cannot Increase Polarization.

second-order stochastically dominates $f^{\lambda'}$; and (d) as $\lambda \downarrow 0$, f^{λ} converges weakly to the degenerate measure granting all weight to μ.

Notice that there is nothing in the definition that requires a squeeze to be applied to symmetric unimodal densities with compact support. In principle, a squeeze as defined could be applied to any density. However, the axioms to be placed below acquire additional cogency when limited to such densities.

Statement of the Axioms. We will impose four axioms on the polarization measure.

Axiom 1. If a distribution is composed of a *single* basic density, then a squeeze of that density cannot increase polarization.

Axiom 1 is self-evident. A squeeze, as defined here, corresponds to a *global* compression of any basic density. If only one of these makes up the distribution (see Figure 1), then the distribution is globally compressed and we must associate this with no higher polarization. Viewed in the context of our background model, however, it is clear that Axiom 1 is going to generate some interesting restrictions. This is because a squeeze creates a reduction in inter-individual alienation but also serves to raise identification for a positive measure of agents — those located "centrally" in the distribution. The implied restriction is, then, that the latter's positive impact on polarization must be counterbalanced by the former's negative impact.

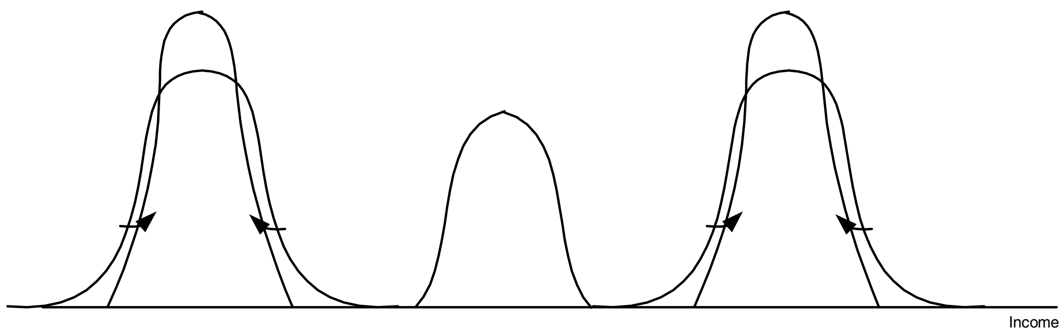

Figure 2: —A Double Squeeze Cannot Lower Polarization.

Our next axiom considers an initial situation (see Figure 2) composed of three disjoint densities all sharing the same root. The situation is completely symmetric, with densities 1 and 3 having the same total population and with density 2 exactly midway between densities 1 and 3.

Axiom 2. If a symmetric distribution is composed of three basic densities with the same root and mutually disjoint supports, then a symmetric squeeze of the *side* densities cannot reduce polarization.

In some sense, this is the defining axiom of polarization, and may be used to motivate the concept. Notice that this axiom argues that a particular "local" squeeze (as opposed to the "global" squeeze of the entire distribution in Axiom 1) must not bring down polarization. At this stage there is an explicit departure from inequality measurement.

Our third axiom considers a symmetric distribution composed of *four* basic densities, once again all sharing the same root.

Axiom 3. Consider a symmetric distribution composed of four basic densities with the same root and mutually disjoint supports, as in Figure 3. Slide the two middle densities to the side as shown (keeping all supports disjoint). Then polarization must go up.

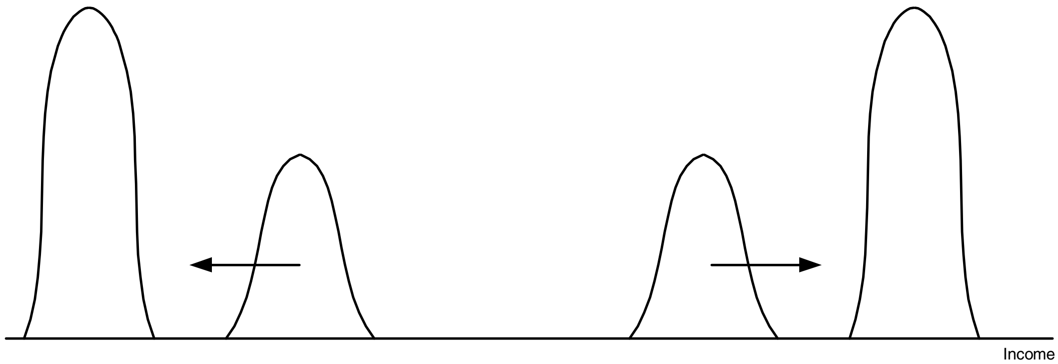

FIGURE 3: —A "Symmetric Outward Slide" Must Raise Polarization.

Our final axiom is a simple population-invariance principle. It states that if one situation exhibits greater polarization than another, it must continue to do so when populations in both situations are scaled up or down by the same amount, leaving all (relative) distributions unchanged.

AXIOM 4. If $P(F) \geq P(G)$ and $p > 0$, then $P(pF) \geq P(pG)$, where pF and pG represent (identical) population scalings of F and G respectively.

2.3. *Characterization Theorem*

THEOREM **1.** *A measure P, as described in (1), satisfies Axioms 1–4 if and only if it is proportional to*

(3) $$P_\alpha(f) \equiv \int\int f(x)^{1+\alpha} f(y)|y-x|dydx,$$

where $\alpha \in [0.25, 1]$.

2.4. *Discussion*

Several aspects of this theorem require extended discussion.

Scaling. Theorem 1 states that a measure of polarization satisfying the preceding four axioms has to be *proportional* to the measure we have characterized. We may wish to exploit this degree of freedom to make the polarization measure scale-free. Homogeneity of degree zero can be achieved, if desired, by multiplying $P_\alpha(F)$ by $\mu^{\alpha-1}$, where μ is mean income. It is easy to see that this procedure is equivalent to one in which all incomes are normalized by their mean, and (3) is subsequently applied.

Importance of the IA Structure. The theorem represents a particularly sharp characterization of the class of polarization measures that satisfy *both* the axioms we have imposed *and* the IA structure. It must be emphasized that both these factors play a role in pinning down our functional form. In fact, it can be checked that several other measures of polarization satisfy Axioms 1–4, though we omit this discussion for the sake of brevity. The IA framework is, therefore, an essential part of the argument.

Partial Ordering. At the same time, and despite the sharpness of the functional form, notice that we do *not* obtain a complete ordering for polarization, nor do we attempt to do this.[9] A range of values of α is entertained in the theorem. The union of the complete orderings generated by each value gives us a partial order for polarization. Pinning down this order completely is an open question.

Identification. A full behavioral foundation for the identification postulate is not within the scope of this paper. However, we make two remarks on the particular specification used here.

First, our axioms imply that identification increases with group size. A well-known problem of collective action (due to Pareto (1906) and Olson (1965)) suggests, however, that smaller groups may sometimes be more effective than larger groups in securing their ends. This argument has been explored by several authors, but perhaps most relevant to the current discussion is Esteban and Ray (2001), which shows that if social conflict arises over the provision of *public goods* (or even if the good is partially private but the cost function for the supply of lobbying resources has sufficient curvature[10]), then larger

[9] Indeed, it is possible to impose additional requirements (along the lines explored by ER, for instance) to place narrower bounds on α. But we do not consider this necessarily desirable. For instance, the upper value $\alpha = 1$ has the property that all λ-squeezes of any distribution leave polarization unchanged. We do not feel that a satisfactory measure *must* possess this feature. This is the reason we are more comfortable with a possible range of acceptable values for α.

[10] Esteban and Ray prove that if the cost function has enough curvature so that it is "at least" quadratic, then larger groups are more effective even if the conflict is over purely private goods.

groups are more effective *in the aggregate*, even though each individual in such groups may be less active owing to the free-rider problem. This finding is consistent with our implication that identification increases with group size.

Second, we remark on our choice of basing identification on the point density. We may more generally suppose that individuals possess a "window of identification" as in ER, section 4. Individuals within this window would be considered "similar" — possibly with weights decreasing with the distance — and would contribute to a sense of group identity. At the same time, individuals would feel alienated only from those outside the window. Thus, broadening one's window of identification has two effects. First, it includes more neighbors when computing one's sense of identification. Second, it reduces one's sense of distance with respect to aliens — because the width of the identification window affects the "starting point" for alienation.

These two effects can be simultaneously captured in our seemingly narrower model. Let t be some parameter representing the "breadth" in identification. Suppose that this means that each individual x will consider an individual with income y to be at the point $(1-t)x + ty$. [Thus t is inversely proportional to "breadth".] The "perceived density" of y from the vantage point of an individual located at x is then

$$\frac{1}{t} f\left(\frac{y-(1-t)x}{t}\right)$$

so that if $t < 1$, the sense of identification is generally heightened (simply set $x = y$ above). Thus a small value of t stands for greater identification.

It can be easily shown that the polarization measure resulting from this extended notion of identification is proportional to our measure by the factor $t^{1-\alpha}$. Therefore, broadening the sense of identification simply amounts to a re-scaling of the measure defined for the limit case in which one is identified with individuals having exactly the same income.

It is also possible to directly base identification on the average density over a non-degenerate window. It can be shown that when our polarization measure is rewritten to incorporate this notion of identification, it converges precisely to the measure in Theorem 1 as the size of the window converges to zero. Thus an alternative view of point-identification is that it is a robust approximation to "narrow" identification windows.

Asymmetric Alienation. In ER we already pointed out that in some environments our implicit hypothesis of a symmetric sense of alienation might not be appropriate. It can be argued that while individuals may feel alienated with respect to those with higher income or wealth, such sentiments need not be reciprocated. For the extreme case of purely one-sided alienation the appropriate extension would be

$$P_\alpha(f) \equiv \int f(x)^{1+\alpha} \int_x f(y)(y-x)dydx.$$

[This is not to say that we have axiomatized such an extension.]

This approach would create a change in the polarization ordering, and depending on the context, it may be a change worth exploring further. The main difference is that (relative to the symmetric case) larger humps or spikes at the lower end of the wealth distribution will be given more weight. In particular, maximal polarization would not be achieved at some symmetric bimodal distribution but at some bimodal distribution that exhibits a larger (local) mode at the bottom of the distribution. This issue is discussed in more detail in ER.

Remarks on the Proof, and the Derived Bounds on α. The proof of Theorem 1 is long and involved, so a brief roadmap may be useful here. The first half of the proof shows that our axioms imply (3), along with the asserted bounds on α. We begin by noting that the function T must be (weakly) concave in alienation (Lemmas 1 and 2). Axiom 2 yields this. Yet by Lemmas 3 and 4 (which centrally employ Axiom 3), T must be (weakly) convex as well. These two assertions must imply that T is linear in alienation, and so is of the form $T(i,a) = \phi(i)a$ for some function ϕ. (Lemma 4 again). Lemma 5 completes the derivation of our functional form by using the population invariance principle (Axiom 4) to argue that ϕ must exhibit constant elasticity.

Our measure bears an interesting resemblance to the Gini coefficient. Indeed, if $\alpha = 0$, the measure *is* the Gini coefficient. However, our arguments ensure that not only is $\alpha > 0$, it cannot go below some uniformly positive lower bound, which happens to be 0.25. Where, in the axioms and in the IA structure, does such a bound lurk? To appreciate this, consider Axiom 2, which refers to a double-squeeze of two "side" basic densities. Such squeezes bring down internal alienations in each component density. Yet the axiom demands that overall polarization not fall. It follows, therefore, that the increased identifications created by the squeeze must outweigh the decreased within-component alienation. This restricts α. It cannot be too low.

By a similar token, α cannot be too high either. The bite here comes from Axiom 1, which decrees that a single squeeze (in an environment where there is just one basic component) cannot increase polarization. Once again, alienation comes down and some identifications go up (as the single squeeze occurs), but this time we want the decline in alienation to dominate the proceedings. This is tantamount to an upper bound on α.[11]

The above arguments are made using Lemmas 6 and 7, which also begin the proof that the axioms are *implied* by our class of measures. The various steps for this direction of the proof, which essentially consist in verifying the axioms, are completed in Lemmas 8 through 11.

The approach to our characterization bears a superficial similarity to ER. Actually, the axioms *are* similar in spirit, dealing as they do in each case with issues of identification and alienation. However, their specific structure is fundamentally different. This is because our axioms strongly exploit the density structure of the model (in ER there are only discrete groupings). In turn, this creates basic differences in the method of proof. It is comforting that the two approaches yield the same functional characterization in the end, albeit with different numerical restrictions on the value of α.

2.5. *Comparing Distributions*

The fundamental hypothesis underlying all of our analysis is that polarization is driven by the interplay of two forces: identification with one's own group and alienation vis-a-vis others. Our axioms yield a particular functional form to the interaction between these two forces. When comparing two distributions, which should we expect to display the greater polarization? Our informal answer is that this should depend on the separate contributions of alienation and identification and on their joint co-movement. Increased alienation is associated with an increase in income distances. Increased identification

[11] One might ask: why do the arguments in this paragraph and the one just before lead to "compatible" thresholds for α? The reason is this: in the double-squeeze, there are cross-group alienations as well which permit a given increase in identification to have a stronger impact on polarization. Therefore the required threshold on α is smaller in this case.

would manifest itself in a sharper definition of groups, *i.e.*, the already highly populated points in the distribution becoming even more populated at the expense of the less populated. Such a change would produce an increase in the variability of the density over the support of the distribution. Finally, when taken jointly, these effects may reinforce each other in the sense that alienation may be highest at the incomes that have experienced an increase in identification, or they may counterbalance each other.

To be sure, it is not possible to move these three factors around independently. After all, one density describes the income distribution and the three factors we have mentioned are byproducts of that density. Nevertheless, thinking in this way develops some intuition for polarization, which we will try and put to use in Section 3.2.

To pursue this line of reasoning, first normalize all incomes by their mean to make the results scale free. Fix a particular value of α, as given by Theorem 1. [More on this parameter below.] The *α-identification* at income y, denoted by $\iota_\alpha(y)$, is measured by $f(y)^\alpha$. Hence, the *average* α-identification $\bar{\iota}$ is defined by

$$(4) \quad \bar{\iota}_\alpha \equiv \int f(y)^\alpha dF(y) = \int f(y)^{1+\alpha} dy.$$

The alienation between two individuals with incomes y and x is given by $|y-x|$. Therefore, the overall alienation felt by an individual with income y, $a(y)$, is

$$(5) \quad a(y) = \int |y-x| dF(x)$$

and the average alienation $\bar{a}$ is

$$(6) \quad \bar{a} = \int a(y) dF(y) = \int \int |y-x| dF(x) dF(y).$$

[Notice that $\bar{a}$ is twice the Gini coefficient.] Now conduct a completely routine exercise. Define ρ as the *normalized covariance* between identification and alienation: $\rho \equiv \mathrm{cov}_{\iota_\alpha,a}/\bar{\iota}_\alpha \bar{a}$. Then

$$\begin{aligned}
\rho \equiv \frac{\mathrm{cov}_{\iota_\alpha,a}}{\bar{\iota}_\alpha \bar{a}} &= \frac{1}{\bar{\iota}_\alpha \bar{a}} \int [\iota_\alpha(y) - \bar{\iota}_\alpha][a(y) - \bar{a}] f(y) dy \\
&= \frac{1}{\bar{\iota}_\alpha \bar{a}} \left[\int f(y)^{1+\alpha} a(y) dy - \bar{a}\bar{\iota}_\alpha \right] \\
&= \frac{P_\alpha(f)}{\bar{\iota}_\alpha \bar{a}} - 1,
\end{aligned}$$

so that

(7) $P_\alpha(f) = \bar{a}\bar{\iota}_\alpha \left[1+\rho\right].$

This is a more precise statement of the informal idea expressed at the start of this section.

There is one dimension, however, along which this decomposition lacks intuition. It is that α unavoidably enters into it: we make this explicit by using the term α-identification (though we will resort to "identification" when there is little risk of confusion). This sort of identification is not intrinsic to the density. Yet the formula itself is useful, for it tells us that — all other things being equal — greater variation, "spikiness", or multimodality in the density is likely to translate into greater polarization for that density, this effect making itself felt more strongly when α is larger. The reason is simple: the main ingredient for α-identification is the function $x^{1+\alpha}$ (see (4)), which is a strictly convex function of x.

The connection with spikiness or multimodality ties in with our graphical intuitions regarding polarization. We reiterate, however, that this is only one factor of several, and that often it may not be possible to change this factor in the direction of higher polarization without infringing the *ceteris paribus* qualification. For instance, if a unimodal density is altered by the introduction of two or more local modes, such multimodality *per se* may not bring higher polarization with it. This is because the existence of several modes may *also* bring average alienation down. In particular, a highly skewed distribution with a single mode may still exhibit greater polarization relative to other bimodal distributions. Nevertheless, the connection with "variability" may be helpful in some situations, and we will invoke it in the empirical discussion of Section 3.2. Indeed, in unimodal situations (which present the most subtle problems as far as polarization is concerned), these factors can act as guides to simple visual inspection.

3. Estimation and Illustration

3.1. *Estimation and Statistical Inference*

We now turn to estimation issues regarding $P_\alpha(F)$ and associated questions of statistical inference. The details of this discussion have been omitted to economize on space, but may be found in the working paper version of this paper (Duclos, Esteban and Ray (2003)). First note that for every distribution function F with associated density f and mean μ, we have that

$$(8) \quad P_\alpha(F) = \int_y f(y)^\alpha a(y)\, dF(y),$$

with $a(y) \equiv \mu + y\,(2F(y) - 1) - 2\int_{-\infty}^{y} x dF(x)$. Suppose, then, that we wish to estimate $P_\alpha(F)$ using a random sample of n iid observations of income y_i, $i = 1, ..., n$, drawn from the distribution $F(y)$ and ordered such that $y_1 \leq y_2 \leq ... \leq y_n$. A natural estimator of $P_\alpha(F)$ is

$$(9) \quad P_\alpha(\widehat{F}) = n^{-1} \sum_{i=1}^{n} \widehat{f}(y_i)^\alpha \widehat{a}(y_i),$$

where $\widehat{a}(y_i)$ is given as

$$(10) \quad \widehat{a}(y_i) = \widehat{\mu} + y_i\left(n^{-1}(2i-1) - 1\right) - n^{-1}\left(2\sum_{j=1}^{i-1} y_j + y_i\right),$$

$\widehat{\mu}$ is the sample mean, and where $\widehat{f}(y_i)^\alpha$ is estimated non-parametrically using kernel estimation procedures[12]. These procedures use a symmetric kernel function $K(u)$, defined such that $\int_{-\infty}^{\infty} K(u)du = 1$ and $K(u) \geq 0$ — a Gaussian kernel is used in the illustration. The estimator $\widehat{f}(y)$ is then defined as $\widehat{f}(y) \equiv n^{-1}\sum_{i=1}^{n} K_h\,(y - y_i)$, with $K_h\,(z) \equiv h^{-1}K\,(z/h)$ and h being a bandwidth parameter. A common technique to select an "optimal" bandwidth h^* is to minimize the mean square error (MSE) of the estimator, given a sample of size n. A "rule-of-thumb" formula that can be used to do this in our context is approximately given by:

$$(11) \quad h^* \cong 4.7\, n^{-0.5}\, \sigma\alpha^{0.1}.$$

Easily computed, this formula works well with the normal distribution since it is then never farther than 5% from the h^* that truly minimizes the MSE. For skewness larger

[12] The literature on kernel density estimation is large – see for instance Silverman (1986), Härdle (1990) and Pagan and Ullah (1999) for an introduction to it.

than about 6, a more robust — though more cumbersome — approximate formula for the computation of h^* is given by

$$(12)\quad h^* \cong n^{-0.5} IQ \frac{(3.76 + 14.7\sigma_{ln})}{(1 + 1.09 \cdot 10^{-4}\sigma_{ln})^{(7268+15323\alpha)}},$$

where IQ is the interquartile and σ_{ln} is the variance of the logarithms of income.

It can also be shown (under certain mild regularity conditions) that $n^{0.5}\left(P_\alpha(\widehat{F}) - P_\alpha(F)\right)$ has an asymptotic limiting normal distribution $N(0, V_\alpha)$, with

$$(13)\quad V_\alpha = \underset{f(y)}{\text{var}}\left((1+\alpha)f(y)^\alpha a(y) + y\int f(x)^\alpha dF(x) + 2\int_y^\infty (x-y)\, f(x)^\alpha dF(x)\right).$$

This result is distribution-free in the sense that everything in the above can be estimated consistently *without* having to specify the population distribution from which the sample is drawn.

3.2. *An Illustration*

We illustrate the above results with data drawn from the Luxembourg Income Study (LIS) data sets[13] on 21 countries for each of Wave 3 (1989–1992) and Wave 4 (1994–1997). Countries, survey years and abbreviations are listed in Table I. [All figures and tables for this section are located at the end of the paper.] We use household disposable income (*i.e.*, post-tax-and-transfer income) normalized by an adult-equivalence scale defined as $s^{0.5}$, where s is household size. Observations with negative incomes are removed as well as those with incomes exceeding 50 times the average (this affects less than 1% of all samples). Household observations are weighted by the LIS sample weights times the number of persons in the household. As discussed in Section 2.4, the usual homogeneity-of-degree-zero property is imposed throughout by multiplying the indices $P_\alpha(F)$ by $\mu^{\alpha-1}$ or equivalently by normalizing all incomes by their mean. For ease of comparison, all indices are divided by 2, so that $P_{\alpha=0}(F)$ is the usual Gini coefficient.

Tables II and III show estimates of the Gini (P_0) and four polarization indices (P_α for $\alpha = 0.25, 0.5, 0.75, 1$) in 21 countries for each of the two waves, along with their asymptotic standard deviations. The polarization indices are typically rather precisely

[13] See `http://lissy.ceps.lu` for detailed information on the structure of these data.

estimated, with often only the third decimal of the estimators being subject to sampling variability. Using a conventional test size of 5%, it can be checked that around 90% of the possible cross-country comparisons are statistically significant, whatever the value of α. Tables II and III also show the country rankings, with a high rank corresponding to a relatively large value of the relevant index, and with countries displayed by their order in the Gini ranking.

Polarization Behaves Differently from Inequality. Observe first that P_0 and $P_{0.25}$ induce very similar rankings. But considerable differences arise between P_0 and P_1. For instance, for Wave 3, the Czech Republic has the lowest Gini index of all countries, but ranks 11 in terms of P_1. Conversely, Canada, Australia and the United States exhibit high Gini inequality, but relatively low "P_1-polarization". The correlation across country rankings for different α's clearly falls as the distance between the α's increases. The lowest Pearson correlation of all — 0.6753 — is the correlation between the Gini index and P_1 in Wave 3. Clearly, polarization and inequality are naturally correlated, but they are also *empirically* distinct in this dataset. Moreover, the extent to which inequality comparisons resemble polarization comparisons depend on the parameter α, which essentially captures the power of the identification effect.

Alienation and Identification. Recall the decomposition exercise carried out in Section 2.5, in which we obtained (7), reproduced here for convenience:

$$P_\alpha = \bar{a}\bar{\iota}_\alpha \left[1 + \rho\right].$$

Table IV summarizes the relevant statistics for all Wave 3 countries, decomposing polarization as the product of average alienation, average identification and (one plus) the normalized covariance between the two. Consider $\alpha = 1$. Note that the bulk of cross-country variation in polarization stems from significant variation in average identification as well as in average alienation. In contrast, the covariance between the two does not exhibit similar variation across countries. Some countries (Finland, Sweden and Denmark) rank low both in terms in inequality and polarization, due in large part to low average alienation. Some countries, most strikingly Russia, Mexico, and the UK rank consistently high both in terms in inequality and polarization — even though average identification

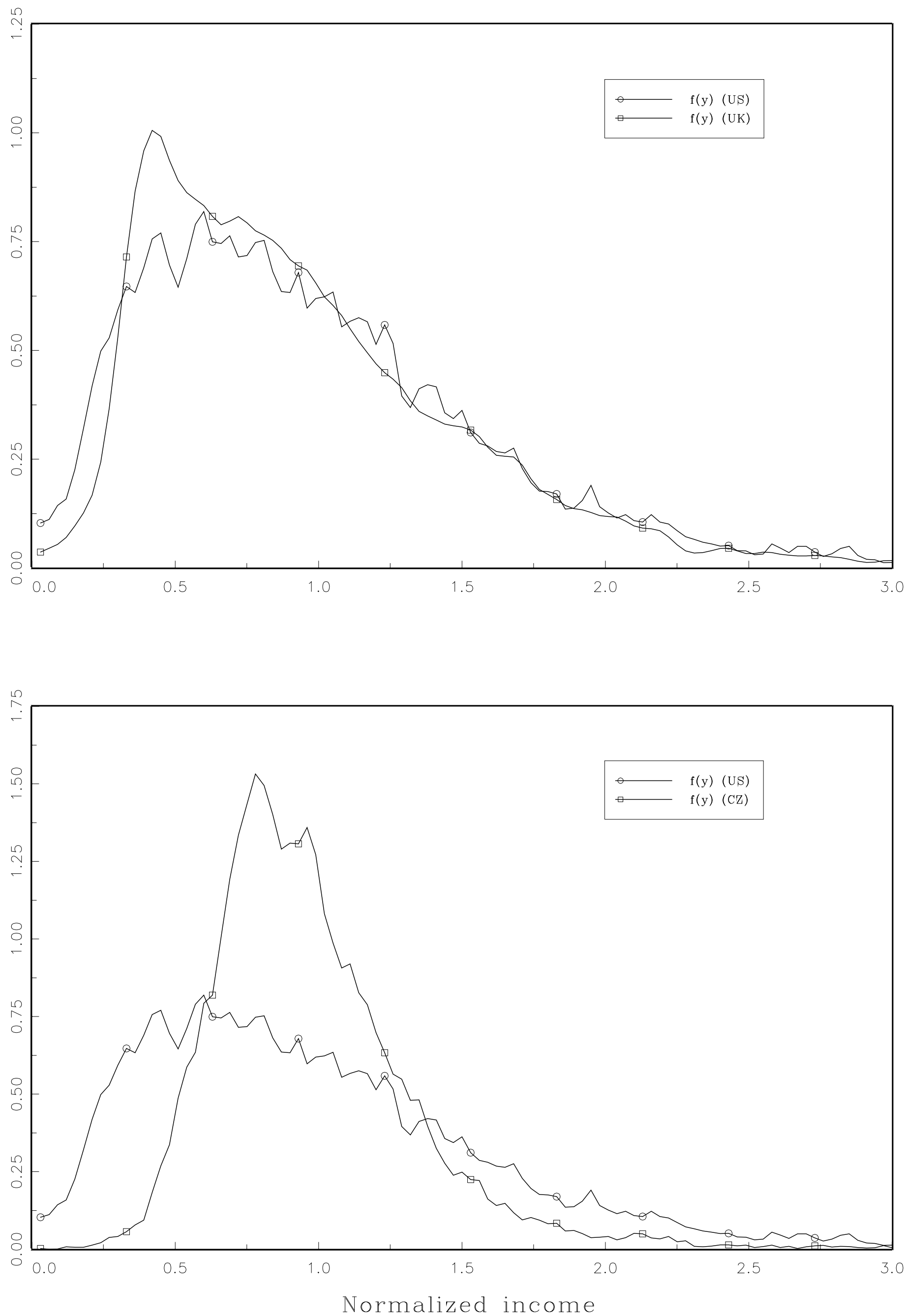

FIGURE 4: —Estimated Densities for the U.S., U.K. and Czech Republic, Wave 3

for the three countries is among the lowest of all. Average alienation is very high in these countries. Yet other countries show low inequality but relatively high polarization, while others exhibit the reverse relative rankings.

While maintaining the same average alienation, the UK density exhibits higher variability than its US counterpart — see the upper panel in Figure 4. The US distribution shows a remarkably flat density on the interval $[0.25, 1.25]$ of normalized incomes and so has thick tails. In contrast, the UK displays a clear mode at $y = 0.4$ and thinner tails. Because — as already discussed in Section 2.5 — the identification function f^α times the density f is strictly convex in f, the country with the greater variation in identification will exhibit a higher value of average identification, *with the difference growing more pronounced as α increases*. To be sure, variations in identification find their starkest expression when distributions are multimodal, but even without such multimodality, variation is possible.

Remember that variation in identification is only one of several factors. In particular, we do not mean to suggest that the country with the greater variation in identification will invariably exhibit greater polarization as $\alpha \to 1$. For instance, our notion of a squeeze increases the variability of identification, but polarization must fall, by Axiom 1 (this is because alienation falls too with the squeeze). See the discussion at the end of Section 2.5 for another illustration of this point.

Sensitivity to α. As α increases from 0.25 to 1, the cross-country variation in the value of average α-identification goes up. This increase in cross-country variability produces frequent "crossings" in the ranking of countries by polarization. Such crossings can occur at very low values of α (below 0.25) so that for *all* $\alpha \in [0.25, 1]$ the polarization ranking opposes the inequality ranking. This is the case (for Wave 3) for Belgium-Sweden, Italy-Canada and Israel-Australia. Crossings could — and do — occur for intermediate values of $\alpha \in [0.25, 1]$. To be sure, they may not occur for any $\alpha \leq 1$, thus causing the polarization ordering to coincide with the inequality ordering. This is indeed a most frequent case for pairwise comparisons in Wave 3. Finally, in Wave 4 we also observe "double crossings" in the cases of Canada-France and Australia-Poland. In both cases the first country starts with higher inequality, P_0, followed by a lower value of $P_{0.25}$, but later returning to higher values for larger values of α.

UK inequality is very close to US inequality; for all intents and purposes the two have the same Gini in Wave 3. Indeed, the UK ranks eighteenth and the US nineteenth – this is true for any $\alpha < 0.33$. However, as α increases beyond 0.33 up to 1, the UK retains the nineteenth position, while the US descends to ninth in the rankings. That fall in rankings occurs mostly when α increases beyond 0.75.

The Czech-US densities provide additional visual support to this sensitivity — see the lower panel in Figure 4. Here, the basic inequality comparison is unambiguous: the Czech Republic has lower inequality than the US. But the Czech Republic has a spikier density with greater variation in it. This "shadow of multimodality" kicks in as α is increased, so much so that the Czech Republic is actually deemed equally or more polarized than the US by the time $\alpha = 1$.

Partial Ordering. One might respond to the above observations as follows: our axiomatics do *not* rule out values of α very close to 0.25. Hence, in the strict sense of a partial order we are unable to (empirically) distinguish adequately between inequality and polarization, at least with the dataset at hand. In our opinion this response would be too hasty. Our characterization not only implies a partial ordering, it provides a very clean picture of how that ordering is parameterized, with the parameter α having a definite interpretation. If substantial variations in ranking occur as α increases, this warrants a closer look, and certainly shows — empirically — how "large" subsets of polarization indices work very differently from the Gini inequality index.[14]

4. FINAL REMARKS, AND A PROPOSED EXTENSION

In this paper we present and characterize a class of measures for income polarization, based on what we call the identification-alienation structure. Our approach is fundamentally based on the view that inter-personal alienation fuels a polarized society, as does inequality. Our departure from inequality measurement lies in the notion that such alienation must also be complemented by a sense of identification. This combination of the two forces generates a class of measures that are sensitive (in the same direction) to both elements of inequality and equality, depending on where these changes are located in the

[14] One would expect these distinctions to magnify even further for distributions that are not unimodal (unfortunately, this exploration is not permitted by our dataset). For instance, one might use our measures to explore the "twin-peaks" property identified by Quah (1996) for the world distribution of income. But this is the subject of future research.

overall distribution.

Our characterization, and the alternative decomposition presented in (7), permit us to describe the measure very simply: for any income distribution, polarization is the product of average alienation, average identification, and (one plus) the mean-normalized covariance between these two variables. We also discuss estimation issues for our measures as well as associated questions of statistical inference.

We wish to close this paper with some remarks on what we see to be the main conceptual task ahead. Our analysis generates a certain structure for identification and alienation functions *in the special case in which both identification and alienation are based on the same characteristic*. This characteristic can be income or wealth. In principle it could be any measurable feature with a well-defined ordering. The key restriction, however, is that whatever we choose the salient characteristic for identification to be, inter-group alienation has to be driven by the very same characteristic. This seems obvious in the cases of income or wealth. Yet, for some relevant social characteristics this might not be a natural assumption. Think of the case of ethnic polarization. It may or may not seem appropriate here to base inter-ethnic alienation as *only* depending on some suitably defined "ethnicity distance". In the cases of socially based group identification we find it more compelling to adopt a multi-dimensional approach to polarization, permitting alienation to depend on characteristics other than the one that defines group identity. In this proposed extension, we liberally transplant our findings to the case of social polarization, but with no further axiomatic reasoning. In our opinion, such reasoning is an important subject of future research.

Suppose, then, that there are M "social groups", based on region, kin, ethnicity, religion... Let n_j be the number of individuals in group j, with overall population normalized to one. Let F_j describe the distribution of income in group j (with f_j the accompanying density), unnormalized by group population. One may now entertain a variety of "social polarization measures".

4.1. *Pure Social Polarization*

Consider, first, the case of "pure social polarization", in which income plays no role. Assume that each person is "fully" identified with every other member of his group. Likewise, the alienation function takes on values that are specific to group pairs and have no reference to income. For each pair of groups j and k denote this value by Δ_{jk}. Then a natural transplant of (3) yields the measure

$$(14)\quad P_s(\mathbf{F}) = \sum_{j=1}^{M}\sum_{k=1}^{M} n_j^{\alpha} n_k \Delta_{jk}.$$

Even this sort of specification may be too general in some interesting instances in which individuals are interested only in the dichotomous perception Us/They. In particular, in these instances, individuals are not interested in differentiating between the different opposing groups. Perhaps the simplest instance of this is a pure contest (Esteban and Ray [1999]), which yields the variant[15]

$$(15)\quad \tilde{P}_s(\mathbf{F}) = \sum_{j=1}^{M} n_j^{\alpha}(1-n_j).$$

4.2. *Hybrids*

Once the two extremes — pure income polarization and pure social polarization — are identified, we may easily consider several hybrids. As examples, consider the case in which notions of identification are mediated not just by group membership but by income similarities as well, while the antagonism equation remains untouched. Then we get what one might call *social polarization with income-mediated identification*:

$$(16)\quad P_s(\mathbf{F}) = \sum_{j=1}^{M}(1-n_j)\int_x f_j(x)^{\alpha} dF_j(x).$$

One could expand (or contract) the importance of income further, while still staying away from the extremes. For instance, suppose that — in addition to the income-mediation of group identity — alienation is also income-mediated (for alienation, two individuals must belong to different groups *and* have different incomes). Now groups have only a

[15] See Reynal-Querol [2002] for a similar analysis. D'Ambrosio and Wolff [2001] also consider a measure of this type but with income distances across groups explicitly considered.

demarcating role — they are necessary (but not sufficient) for identity, and they are necessary (but not sufficient) for alienation. The resulting measure would look like this:

$$(17) \quad P^*(\mathbf{F}) = \sum_{j=1}^{M} \sum_{k \neq j} \int_x \int_y f_j(x)^\alpha |x - y| dF_j(x) dF_k(y).$$

Note that we do not intend to suggest that other special cases or hybrids are not possible, or that they are less important. The discussion here is only to show that social and economic considerations can be profitably combined in the measurement of polarization. Indeed, it is conceivable that such measures will perform better than the more commonly used fragmentation measures in the analysis of social conflict. But a full exploration of this last theme must await a future paper.

Département d'économique, Pavillon J.-A.-DeSève, Université Laval, Québec, Canada G1K 7P4; jyves@ecn.ulaval.ca

Instituto de Análisis Económico, Universidad Autónoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; Joan.Esteban@uab.es

and

Department of Economics, New York University, 269 Mercer Street, New York, NY 10003, U.S.A.; debraj.ray@nyu.edu, http://www.econ.nyu.edu/user/debraj/

APPENDIX

Proof of Theorem 1. First, we show that axioms 1–4 imply (3). The lemma below follows from Jensen's inequality; proof omitted.

LEMMA **1.** *Let g be a continuous real-valued function defined on $\mathbb{R}$ such that for all $x > 0$ and all δ with $0 < \delta < x$,*

$$(18) \quad g(x) \geq \frac{1}{2\delta}\int_{x-\delta}^{x+\delta} g(y)dy.$$

Then g must be a concave function.

In what follows, remember that our measure only considers income *differences* across people, so that we may slide any distribution to left or right as we please.

LEMMA **2.** *The function T must be concave in a for every $i > 0$.*

Proof. Fix $x > 0$, some $i > 0$, and $\delta \in (0, x)$. Consider three basic densities as in Axiom 2 (see Figure 1) but specialize as shown in Figure 5; each is a transform of a uniform basic density. The bases are centered at $-x$, 0 and x. The side densities are of width 2δ and height h, and the middle density is of width 2ϵ and height i. We shall vary ϵ and h but to make sure that Axiom 2 applies, we choose $\epsilon > 0$ such that $\delta + \epsilon < x$. A λ-squeeze of the side densities simply contracts their base width to $2\lambda\delta$, while the height is raised to h/λ. For each λ, decompose the measure (1) into five components. (a) The "internal polarization" P_m of the middle rectangle. This component doesn't vary with λ so there will be no need to explicitly calculate it. (b) The "internal polarization" P_s of each side rectangle. (c) Total effective antagonism, A_{ms} felt by inhabitants of the middle towards each side density. (d) Total effective antagonism A_{sm} felt by inhabitants of each side towards the middle. (e) Total effective antagonism A_{ss} felt by inhabitants of one side towards the other side. Each of these last four terms appear twice, so that (writing everything as a function of λ),

$$(19) \quad P(\lambda) = P_m + 2P_s(\lambda) + 2A_{ms}(\lambda) + 2A_{sm}(\lambda) + 2A_{ss}(\lambda),$$

Now we compute the terms on the right hand side of (19). First,

$$P_s(\lambda) = \frac{1}{\lambda^2}\int_{x-\lambda\delta}^{x+\lambda\delta}\int_{x-\lambda\delta}^{x+\lambda\delta} T(h/\lambda, |b' - b|)h^2 db' db,$$

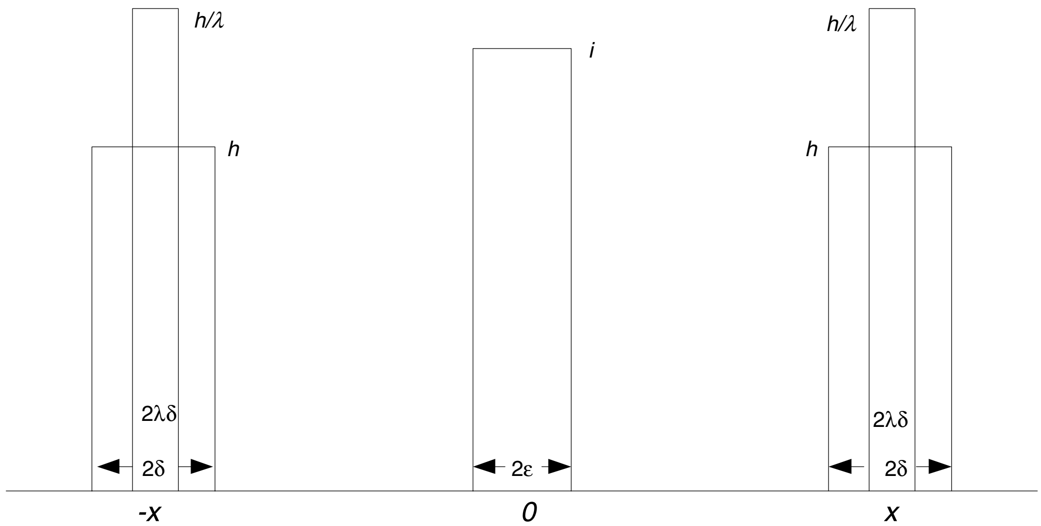

FIGURE 5:

where (here and in all subsequent cases) b will stand for the "origin" income (to which the identification is applied) and b' the "destination income" (towards which the antagonism is felt). Next,

$$A_{ms}(\lambda) = \frac{1}{\lambda} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda\delta}^{x+\lambda\delta} T(i, b'-b) ihdb'db.$$

Third,

$$A_{sm}(\lambda) = \frac{1}{\lambda} \int_{x-\lambda\delta}^{x+\lambda\delta} \int_{-\epsilon}^{\epsilon} T(h/\lambda.b-b') hidb'db,$$

And finally,

$$A_{ss}(\lambda) = \frac{1}{\lambda^2} \int_{-x-\lambda\delta}^{-x+\lambda\delta} \int_{x-\lambda\delta}^{x+\lambda\delta} T(h/\lambda, b'-b) h^2 db'db.$$

The axiom requires that $P(\lambda) \geq P(1)$. Equivalently, we require that $[P(\lambda)-P(1)]/2h \geq 0$ for all h, which implies in particular that

$$(20) \quad \liminf_{h \to 0} \frac{P(\lambda) - P(1)}{2h} \geq 0.$$

If we divide through by h in the individual components calculated above and then send h to 0, it is easy to see that the only term that remains is A_{ms}. Formally, (20) and the calculations above must jointly imply that

$$(21) \quad \frac{1}{\lambda} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda\delta}^{x+\lambda\delta} T(i, b'-b) db'db \geq \int_{-\epsilon}^{\epsilon} \int_{x-\delta}^{x+\delta} T(i, b'-b) db'db,$$

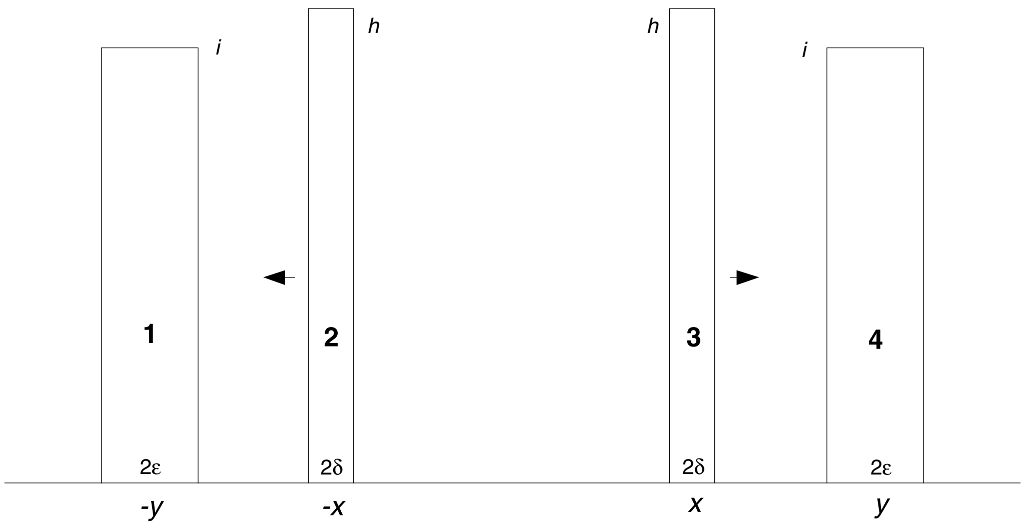

FIGURE 6:

and this must be true for all $\lambda \in (0,1)$ as well as all $\epsilon \in (0, x-\delta)$. Therefore we may insist on the inequality in (21) holding as $\lambda \to 0$. Performing the necessary calculations, we may conclude that

$$(22) \quad \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon} T(i, x-b)db \geq \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon}\int_{x-\delta}^{x+\delta} T(i, b'-b)db'db$$

for every $\epsilon \in (0, x-\delta)$. Finally, take ϵ to zero in (22). This allows us to deduce that

$$(23) \quad T(i,x) \geq \int_{x-\delta}^{x+\delta} T(i,b')db'.$$

As (23) must hold for every $x > 0$ and every $\delta \in (0,x)$, we may invoke Lemma 1 to conclude that T is concave in x for every $i > 0$.

Q.E.D.

LEMMA **3.** *Let g be a concave, continuous function on $I\!R_+$, with $g(0) = 0$. Suppose that for each a and a' with $a > a' > 0$, there exists $\bar{\Delta} > 0$ such that*

$$(24) \quad g(a+\Delta) - g(a) \geq g(a') - g(a'-\Delta)$$

for all $\Delta \in (0, \bar{\Delta})$. Then g must be linear.

The proof is straightforward and is omitted.

Lemma 4. *There is a continuous function $\phi(i)$ such that $T(i,a) = \phi(i)a$ for all i and a.*

Proof. Fix a and a' with $a > a' > 0$, and $i > 0$. Consider four basic densities as in Axiom 3 (see Figure 3) but specialize as shown in Figure 6; each is a transform of a uniform basic density. The bases are centered at locations $-y$, $-x$, x and y, where $x \equiv (a-a')/2$ and $y \equiv (a+a')/2$. The "inner" densities are of width 2δ and height h, and the "outer" densities are of width 2ϵ and height i. We shall vary different parameters (particularly x) but to ensure disjoint support we assume throughout that $\epsilon < x$ and $\delta + \epsilon < y - x - \bar{\Delta}$ for some $\bar{\Delta} > 0$. Again, decompose the polarization measure (1) into several distinct components. (a) The "internal polarization" of each rectangle j; call it P_j, $j = 1, 2, 3, 4$. These components are unchanged as we change x so there will be no need to calculate them explicitly. (b) Total effective antagonism $A_{jk}(x)$ felt by inhabitants of rectangle j towards rectangle k (we emphasize dependence on the parameter x). Thus total polarization $P(x)$ is given by

$$\begin{aligned} P(x) &= \sum_{j=1}^{4} P_j + \sum_j \sum_{k \neq j} A_{jk}(x) \\ &= \sum_{j=1}^{4} P_j + 2A_{12}(x) + 2A_{13}(x) + 2A_{21}(x) + 2A_{31}(x) + 2A_{23}(x) + 2A_{14}, \end{aligned}$$

where the second equality simply exploits obvious symmetries and A_{14} is noted to be independent of x. Let's compute the terms in this formula that do change with x. We have

$$\begin{aligned} A_{12}(x) &= \int_{-y-\epsilon}^{-y+\epsilon} \int_{-x-\delta}^{-x+\delta} T(i, b'-b) ih db' db, \\ A_{13}(x) &= \int_{-y-\epsilon}^{-y+\epsilon} \int_{x-\delta}^{x+\delta} T(i, b'-b) ih db' db, \\ A_{21}(x) &= \int_{-x-\delta}^{-x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b-b') ih db' db, \\ A_{31}(x) &= \int_{x-\delta}^{x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b-b') ih db' db, \end{aligned}$$

and

$$A_{23}(x) = \int_{-x-\delta}^{-x+\delta} \int_{x-\delta}^{x+\delta} T(h, b-b') h^2 db' db.$$

Now, the axiom requires that $P(x+\Delta) - P(x) \geq 0$. Equivalently, we require that $[P(x+\Delta) - P(1)]/2ih \geq 0$ for all h, which implies in particular that

$$\liminf_{h \to 0} \frac{P(x+\Delta) - P(x)}{2ih} \geq 0.$$

Using this information along with the computations for $P(x)$ and the various $A_{jk}(x)$'s, we see (after some substitution of variables and transposition of terms) that

$$\begin{aligned} &\int_{-y-\epsilon}^{-y+\epsilon} \int_{x-\delta}^{x+\delta} [T(i, b'-b+\Delta) - T(i, b'-b)]\, db'db \\ \geq &\int_{-y-\epsilon}^{-y+\epsilon} \int_{-x-\delta}^{-x+\delta} [T(i, b'-b) - T(i, b'-b-\Delta)]\, db'db, \end{aligned}$$

Dividing through by δ in this expression and then taking δ to zero, we may conclude that

$$\int_{-y-\epsilon}^{-y+\epsilon} [T(i, x-b+\Delta) - T(i, x-b)]\, db \geq \int_{-y-\epsilon}^{-y+\epsilon} [T(i, -x-b) - T(i, -x-b-\Delta)]\, db,$$

and dividing this inequality, in turn, by ϵ and taking ϵ to zero, we see that

$$T(i, a+\Delta) - T(i, a) \geq T(i, a') - T(i, a'-\Delta),$$

where we use the observations that $x+y=a$ and $y-x=a'$. Therefore the conditions of Lemma 3 are satisfied, and $T(i,.)$ must be linear for every $i>0$ since $T(0,a)=0$. That is, there is a function $\phi(i)$ such that $T(i,a) = \phi(i)a$ for every i and a. Given that T is continuous by assumption, the same must be true of ϕ.

Q.E.D.

LEMMA **5.** *$\phi(i)$ must be of the form Ki^α, for constants $(K, \alpha) \gg 0$.*

Proof. As a preliminary step, observe that

(25) $\phi(i) > 0$ whenever $i > 0$.

otherwise Axiom 3 would fail for configurations constructed from rectangular basic densities of equal height i. We first prove that ϕ satisfies the fundamental Cauchy equation

(26) $\phi(p)\phi(p') = \phi(pp')\phi(1)$

for every $(p, p') \gg 0$. To this end, fix p and p' and define $r \equiv pp'$. In what follows, we assume that $p \geq r$.[16] Consider a configuration with two basic densities, both of width 2ϵ, the first centered at 0 and the second centered at 1. The heights are p and h (where $h > 0$ but soon to be made arbitrarily small). A little computation shows that polarization in this case is given by

$$\begin{aligned} P &= ph[\phi(p) + \phi(h)]\{\int_{-\epsilon}^{\epsilon}\int_{1-\epsilon}^{1+\epsilon}(b' - b)db'db\} \\ &\quad + [p^2\phi(p) + h^2\phi(h)]\{\int_{-\epsilon}^{\epsilon}\int_{-\epsilon}^{\epsilon}|b' - b|db'db\} \\ &= 4\epsilon^2 ph[\phi(p) + \phi(h)] + \frac{8\epsilon^3}{3}[p^2\phi(p) + h^2\phi(h)], \end{aligned} \tag{27}$$

where the first equality invokes Lemma 4. Now change the height of the first rectangle to r. Using (25) and $p \geq r$, it is easy to see that for each ϵ, there exists a (unique) height $h(\epsilon)$ for the second rectangle such that the polarizations of the two configurations are equated. Invoking (27), $h(\epsilon)$ is such that

$$\begin{aligned} & ph[\phi(p) + \phi(h)] + \frac{2\epsilon}{3}[p^2\phi(p) + h^2\phi(h)] \\ &= rh(\epsilon)[\phi(r) + \phi(h(\epsilon))] + \frac{2\epsilon}{3}[r^2\phi(r) + h(\epsilon)^2\phi(h(\epsilon))]. \end{aligned} \tag{28}$$

By Axiom 4, it follows that for all $\lambda > 0$,

$$\begin{aligned} & \lambda^2 ph[\phi(\lambda p) + \phi(\lambda h)] + \frac{2\epsilon}{3}[(\lambda p)^2\phi(\lambda p) + (\lambda h)^2\phi(\lambda h)] \\ &= \lambda^2 rh(\epsilon)[\phi(\lambda r) + \phi(\lambda h(\epsilon))] + \frac{2\epsilon}{3}[(\lambda r)^2\phi(\lambda r) + [\lambda h(\epsilon)]^2\phi(\lambda h(\epsilon))]. \end{aligned} \tag{29}$$

Notice that as $\epsilon \downarrow 0$, $h(\epsilon)$ lies in some bounded set. We may therefore extract a convergent subsequence with limit h' as $\epsilon \downarrow 0$. By the continuity of ϕ, we may pass to the limit in (28) and (29) to conclude that

$$ph[\phi(p) + \phi(h)] = rh'[\phi(r) + \phi(h')] \tag{30}$$

and

$$\lambda^2 ph[\phi(\lambda p) + \phi(\lambda h)] = \lambda^2 rh'[\phi(\lambda r) + \phi(\lambda h')]. \tag{31}$$

[16] If $r \geq p$, simply permute p and r in the argument below.

Combining (30) and (31), we see that

$$(32)\quad \frac{\phi(p)+\phi(h)}{\phi(\lambda p)+\phi(\lambda h)} = \frac{\phi(r)+\phi(h')}{\phi(\lambda r)+\phi(\lambda h')}.$$

Taking limits in (32) as $h \to 0$ and noting that $h' \to 0$ as a result (examine (30) to confirm this), we have for all $\lambda > 0$,

$$(33)\quad \frac{\phi(p)}{\phi(\lambda p)} = \frac{\phi(r)}{\phi(\lambda r)}.$$

Put $\lambda = 1/p$ and recall that $r = pp'$. Then (33) yields the required Cauchy equation (26). To complete the proof, recall that ϕ is continuous and that (25) holds. The class of solutions to (26) (that satisfy these additional qualifications) is completely described by $\phi(p) = Kp^\alpha$ for constants $(K, \alpha) \gg 0$ (see, e.g., Aczél [1966, p. 41, Theorem 3]).

Lemmas 4 and 5 together establish "necessity", though it still remains to establish the bounds on α. We shall do so along with our proof of "sufficiency", which we begin now.

Lemma **6.** *Let f be a basic density with mass p and mean μ on support $[a,b]$. Let $m \equiv \mu - a$ and let f^* denote the root of f. Then, if f^λ denotes some λ-squeeze of f,*

(34)

$$P(F^\lambda) = 4kp^{2+\alpha}(m\lambda)^{1-\alpha}\int_0^1 f^*(x)^{1+\alpha}\left\{\int_0^1 f^*(y)(1-y)dy + \int_x^1 f^*(y)(y-x)dy\right\}dx$$

for some constant $k > 0$.

Proof. Recall that a slide of f has no effect on the computations, so we may as well set $a = 0$ and $b = 2m$, where $m = \mu - a$ is now to be interpreted as the mean. Given (3),

$$(35)\quad P(F) = k\int\int f(x)^{1+\alpha} f(y)|y-x|dydx$$

for some $k > 0$. Using the fact that f is symmetric, we can write

$$\begin{aligned} P(F) &= 2k\int_0^m\int_0^{2m} f(x')^{1+\alpha} f(y')|x'-y'|dy'dx' \\ &= 2k\int_0^m f(x')^{1+\alpha}\left\{\int_0^{x'} f(y')(x'-y')dy' + \int_{x'}^m f(y')(y'-x')dy'\right. \\ (36)\qquad &\left. + \int_m^{2m} f(y')(y'-x')dy'\}dx'\right\}. \end{aligned}$$

Examine the very last term in (36). Change variables by setting $z \equiv 2m - y'$, and use symmetry to deduce that

$$\int_m^{2m} f(y')(y' - x')dy' = \int_0^m f(z)(2m - x' - z)dz.$$

Substituting this in (36), and manipulating terms, we obtain

(37) $$P(F) = 4k \int_0^m f(x')^{1+\alpha} \left\{ \int_0^m f(y')(m - y')dy' + \int_{x'}^m f(y')(y' - x')dy' \right\} dx'.$$

Now suppose that f^λ is a λ-squeeze of f. Note that (37) holds just as readily for f^λ as for f. Therefore, using the expression for f given in (2), we see that

$$P(F^\lambda) = 4k\lambda^{-(2+\alpha)} \int_{(1-\lambda)m}^m f\left(\frac{x' - (1-\lambda)m}{\lambda}\right)^{1+\alpha} \left\{ \int_{(1-\lambda)m}^m f\left(\frac{y' - (1-\lambda)m}{\lambda}\right)(m - y')dy' \right.$$
$$\left. + \int_{x'}^m f\left(\frac{y' - (1-\lambda)m}{\lambda}\right)(y' - x')dy' \} dx' \right\}.$$

Perform the change of variables $x'' = \frac{x'-(1-\lambda)m}{\lambda}$ and $y'' = \frac{y'-(1-\lambda)m}{\lambda}$. Then it is easy to see that

$$P(F^\lambda) = 4k\lambda^{1-\alpha} \int_0^m f(x'')^{1+\alpha} \left\{ \int_0^m f(y'')(m - y'')dy'' + \int_{x''}^m f(y'')(y'' - x'')dy'' \right\} dx''.$$

To complete the proof, we must recover the root f^* from f. To this end, first population-scale f to h, where h has mass 1. That is, $f(z) = ph(z)$ for all z. Doing so, we see that

$$P(F^\lambda) = 4kp^{2+\alpha}\lambda^{1-\alpha} \int_0^m h(x'')^{1+\alpha} \left\{ \int_0^m h(y'')(m - y'')dy'' + \int_{x''}^m h(y'')(y'' - x'')dy'' \right\} dx''.$$

Finally, make the change of variables $x = x''/m$ and $y = y''/m$. Noting that $f^*(z) = mh(mz)$, we get (34).

Q.E.D.

Lemma **7.** *Let f and g be two basic densities with disjoint support, with their means separated by distance d, and with population masses p and q respectively. Let f have mean μ on support $[a, b]$. Let $m \equiv \mu - a$ and let f^* denote the root of f. Then for any λ -squeeze f^λ of f,*

(38) $$A(f^\lambda, g) = 2kdp^{1+\alpha}q(m\lambda)^{-\alpha} \int_0^1 f^*(x)^{1+\alpha}dx,$$

where $A(f^\lambda, g)$ denotes the total effective antagonism felt by members of f^λ towards members of g.

Proof. Without loss of generality, let f have support $[0, 2m]$ (with mean m) and g have support $[d, d+2m]$ (where $d \geq 2m$ for disjoint supports). Using (35),

$$\begin{aligned}
A(f,g) &= k\int_0^{2m} f(x)^{1+\alpha}\left[\int_d^{d+2m} g(y)(y-x)dy\right]dx\\
&= k\int_0^{2m} f(x)^{1+\alpha}\left[\int_d^{d+m} g(y)(y-x)dy + \int_{d+m}^{d+2m} g(y)(y-x)dy\right]dx\\
&= k\int_0^{2m} f(x)^{1+\alpha}\left[\int_d^{d+m} g(y)2(m+d-x)dy\right]dx\\
&= kq\int_0^{2m} f(x)^{1+\alpha}(m+d-x)dx\\
&= 2dkq\int_0^{m} f(x)^{1+\alpha}dx,
\end{aligned}$$

where the third equality exploits the symmetry of g,[17] the fourth equality uses the fact that $\int_d^{d+m} g(y) = q/2$, and the final equality uses the symmetry of f.[18] To be sure, this formula applies to any λ -squeeze of f, so that

$$\begin{aligned}
A(f^\lambda, g) &= 2dkq\int_0^{m} f^\lambda(x')^{1+\alpha}dx'\\
&= 2dkq\lambda^{-(1+\alpha)}\int_{(1-\lambda)m}^{m} f\left(\frac{x'-(1-\lambda)m}{\lambda}\right)^{1+\alpha}dx',
\end{aligned}$$

and making the change of variables $x'' = \frac{x'-(1-\lambda)m}{\lambda}$, we may conclude that

$$A(f^\lambda, g) = 2dkq\lambda^{-\alpha}\int_0^{m} f(x'')^{1+\alpha}dx''.$$

To complete the proof, we must recover the root f^* from f. As in the proof of Lemma 6, first population-scale f to h, where h has mass 1. That is, $f(z) = ph(z)$ for all z. Doing so, we see that

$$A(f^\lambda, g) = 2dkp^{1+\alpha}q\lambda^{-\alpha}\int_0^{m} h(x'')^{1+\alpha}dx''.$$

Finally, make the change of variables $x = x''/m$. Noting that $f^*(z) = mh(mz)$, we get (38). ∎

[17] That is, for each $y \in [d, d+m]$, $g(y) = g(d+2m-(y-d)) = g(2d+2m-y)$. Moreover, $[y-x] + [(2d+2m-y)-x] = 2(d+m-x)$.

[18] That is, for each $x \in [0, m]$, $f(x) = f(2m-x)$. Moreover, $[m+d-x] + [m+d-(2m-x)] = 2d$.

LEMMA **8.** *Define, for any root f and $\alpha > 0$,*

$$\psi(f,\alpha) \equiv \frac{\int_0^1 f(x)^{1+\alpha}dx}{\int_0^1 f(x)^{1+\alpha}\left\{\int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy\right\}dx}. \tag{39}$$

Then — for any $\alpha > 0$ — $\psi(f,\alpha)$ attains its minimum value when f is the uniform root, and this minimum value equals 3.

Proof. It will be useful to work with the inverse function

$$\zeta(f,\alpha) \equiv \psi(f,\alpha)^{-1} = \frac{\int_0^1 f(x)^{1+\alpha}\left\{\int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy\right\}dx}{\int_0^1 f(x)^{1+\alpha}dx}.$$

Note that $\zeta(f,\alpha)$ may be viewed as a weighted average of

$$L(x) \equiv \int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy \tag{40}$$

as this expression varies over $x \in [0,1]$, where the "weight" on a particular x is just

$$\frac{f(x)^{1+\alpha}}{\int_0^1 f(z)^{1+\alpha}dz}$$

which integrates over x to 1. Now observe that $L(x)$ is *decreasing* in x. Moreover, by the unimodality of a root, the weights must be nondecreasing in x. It follows that

$$\zeta(f,\alpha) \le \int_0^1 L(x)dx. \tag{41}$$

Now

$$\begin{aligned}
L(x) &= \int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy \\
&= \int_0^1 f(y)(1-x)dy + \int_0^x f(y)(x-y)dy \\
&= \frac{1-x}{2} + \int_0^x f(y)(x-y)dy. \qquad (42)
\end{aligned}$$

Because $f(x)$ is nondecreasing and integrates to 1/2 on $[0,1]$, it must be the case that $\int_0^x f(y)(x-y)dy \le \int_0^x (x-y)/2dy$ for all $x \le 1$. Using this information in (42) and combining it with (41),

$$\begin{aligned}
\zeta(f,\alpha) &\le \int_0^1 \left[\frac{1-x}{2} + \int_0^x \frac{x-y}{2}dy\right]dx \\
&= \int_0^1 \left[\int_0^1 \left[\frac{1-y}{2}\right]dy + \int_x^1 \left[\frac{y-x}{2}\right]dy\right]dx \\
&= \zeta(u,\alpha), \qquad (43)
\end{aligned}$$

where u stands for the uniform root taking constant value $1/2$ on $[0, 2]$. Simple integration reveals that $\zeta(u, \alpha) = 1/3$.

Q.E.D.

LEMMA **9.** *Given that $P(f)$ is of the form (35), Axiom 1 is satisfied if and only if $\alpha \leq 1$.*

Proof. Simply inspect (34).

Q.E.D.

LEMMA **10.** *Given that $P(f)$ is of the form (35), Axiom 2 is satisfied if and only if $\alpha \geq 0.25$.*

Proof. Consider a configuration as given in Axiom 2: a symmetric distribution made out of three basic densities. By symmetry, the side densities must share the same root; call this f^*. Let p denote their (common) population mass and m their (common) difference from their means to their lower support. Likewise, denote the root of the middle density by g^*, by q its population mass, and by n the difference between mean and lower support. As in the proof of Lemma 2, we may decompose the polarization measure (35) into several components. First, there are the "internal polarizations" of the middle density (P_m) and of the two side densities (P_s). Next, there are various subtotals of effective antagonism felt by members of one of the basic densities towards another basic density. Let A_{ms} denote this when the "origin" density is the middle and the "destination" density one of the sides. Likewise, A_{sm} is obtained by permuting origin and destination densities. Finally, denote by A_{ss} the total effective antagonism felt by inhabitants of one side towards the other side. Observe that each of these last four terms appear twice, so that (writing everything as a function of λ), overall polarization is given by

$$(44) \quad P(\lambda) = P_m + 2P_s(\lambda) + 2A_{ms}(\lambda) + 2A_{sm}(\lambda) + 2A_{ss}(\lambda).$$

Compute these terms. For brevity, define for any root h,

$$\psi_1(h, \alpha) \equiv \int_0^1 h(x)^{1+\alpha} \left\{ \int_0^1 h(y)(1-y)dy + \int_x^1 h(y)(y-x)dy \right\} dx$$

and

$$\psi_2(h,\alpha) \equiv \int_0^1 h(x)^{1+\alpha}dx.$$

Now, using Lemmas 6 and 7, we see that

$$P_s(\lambda) = 4kp^{2+\alpha}(m\lambda)^{1-\alpha}\psi_1(f^*,\alpha),$$

while

$$A_{ms}(\lambda) = 2kdq^{1+\alpha}pn^{-\alpha}\psi_2(g^*,\alpha).$$

Moreover,

$$A_{sm}(\lambda) = 2kdp^{1+\alpha}q(m\lambda)^{-\alpha}\psi_2(f^*,\alpha),$$

and

$$A_{ss}(\lambda) = 4kdp^{2+\alpha}(m\lambda)^{-\alpha}\psi_2(f^*,\alpha),$$

(where it should be remembered that the distance between the means of the two side densities is $2d$). Observe from these calculations that $A_{ms}(\lambda)$ is entirely insensitive to λ. Consequently, feeding all the computed terms into (44), we may conclude that

$$P(\lambda) = C\left[2\lambda^{1-\alpha} + \frac{d}{m}\psi(f^*,\alpha)\lambda^{-\alpha}\{\frac{q}{p}+2\}\right] + D,$$

where C and D are positive constants independent of λ, and

$$\psi(f^*,\alpha) = \frac{\psi_2(f^*,\alpha)}{\psi_1(f^*,\alpha)}$$

by construction; see (39) in the statement of Lemma 8. It follows from this expression that for Axiom 2 to hold, it is necessary and sufficient that for *every* three-density configuration of the sort described in that axiom,

$$2\lambda^{1-\alpha} + \frac{d}{m}\psi(f^*,\alpha)\lambda^{-\alpha}\left[\frac{q}{p}+2\right] \tag{45}$$

must be nonincreasing in λ over $(0,1]$. An examination of the expression in (45) quickly shows that a situation in which q is arbitrarily close to zero (relative to p) is a necessary and sufficient test case. By the same logic, one should make d/m as small as possible. The disjoint-support hypothesis of Axiom 2 tells us that this lowest value is 1. So it will be necessary and sufficient to show that for every root f^*,

$$(46) \quad \lambda^{1-\alpha} + \psi(f^*, \alpha)\lambda^{-\alpha}$$

is nonincreasing in λ over $(0, 1]$. For any f^*, it is easy enough to compute the necessary and sufficient bounds on α. Simple differentiation reveals that

$$(1-\alpha)\lambda^{-\alpha} - \alpha\psi(f^*, \alpha)\lambda^{-(1+\alpha)}$$

must be nonnegative for every $\lambda \in (0, 1]$; the necessary and sufficient condition for this is

$$(47) \quad \alpha \geq \frac{1}{1 + \psi(f^*, \alpha)}.$$

Therefore, to find the necessary and sufficient bound on α (uniform over all roots), we need to minimize $\psi(f^*, \alpha)$ by choice of f^*, subject to the condition that f^* be a root. By Lemma 8 , this minimum value is 3. Using this information in (47), we are done. ∎

Lemma **11.** *Given that $P(f)$ is of the form (35), Axiom 3 is satisfied.*

Proof. Consider a symmetric distribution composed of four basic densities, as in the statement of Axiom 3. Number the densities 1, 2, 3 and 4, in the same order displayed in Figure 6. Let x denote the amount of the slide (experienced by the inner densities) in the axiom. For each such x, let $d_{jk}(x)$ denote the (absolute) difference between the means of basic densities j and k. As we have done several times before, we may decompose the polarization of this configuration into several components. First, there is the "internal polarization" of each rectangle j; call it P_j, $j = 1, 2, 3, 4$. [These will stay unchanged with x.] Next, there is the total effective antagonism felt by inhabitants of each basic density towards another; call this $A_{jk}(x)$, where j is the "origin" density and k is the "destination" density. Thus total polarization $P(x)$, again written explicitly as a function of x, is given by

$$P(x) = \sum_{j=1}^{4} P_j + \sum_j \sum_{k \neq j} A_{jk}(x)$$

so that, using symmetry,

$$(48)\quad P(x) - P(0) = 2\{[A_{12}(x) + A_{13}(x)] - [A_{12}(0) + A_{13}(0)]\} + [A_{23}(x) - A_{23}(0)]$$

Now Lemma 7 tells us that for all i and j,

$$A_{ij}(x) = k_{ij}d_{ij}(x),$$

where k_{ij} is a positive constant which is independent of distances across the two basic densities, and in particular is independent of x. Using this information in (48), it is trivial to see that

$$P(x) - P(0) = A_{23}(x) - A_{23}(0) = k_{ij}x > 0,$$

so that Axiom 3 is satisfied. ∎

Given (35), Axiom 4 is trivial to verify. Therefore Lemmas 9, 10 and 11 complete the proof of the theorem.

Q.E.D.

ADDENDUM

This addendum contains details and proofs of the estimation and statistical inference results that are reported at the beginning of Section 3.

Estimating $P_\alpha(F)$

The following rewriting of $P_\alpha(F)$ is useful:

Observation **1.** *For every distribution function F with associated density f and mean μ,*

$$P_\alpha(F) = \int_y f(y)^\alpha a(y)\, dF(y) \equiv \int_y p_\alpha(y)\, dF(y), \tag{49}$$

with $a(y) \equiv \mu + y\,(2F(y) - 1) - 2\mu^*(y)$, where $\mu^*(y) = \int_{-\infty}^{y} x dF(x)$ is a partial mean and where $p_\alpha(y) = f(y)^\alpha a(y)$.

Proof of Observation 1. First note that $|x - y| = x + y - 2\min(x, y)$. Hence, by (3),

$$P_\alpha(f) = \int_x \int_y f(y)^\alpha\,(x + y - 2\min(x, y))\, dF(y)\, dF(x).$$

To prove (49), note that

$$\int_x \int_y x\, f(y)^\alpha\, dF(y)\, dF(x) = \mu \int_y f(y)^\alpha\, dF(y) \tag{50}$$

and that

$$\begin{aligned} &\int_x \int_y f(y)^\alpha \min(x, y)\, dF(y) dF(x) \\ &= \int_x \int_{y=-\infty}^{y=x} y f(y)^\alpha dF(y) dF(x) + \int_x \int_{y=x}^{\infty} x f(y)^\alpha dF(y) dF(x). \end{aligned} \tag{51}$$

The first term in (51) can be integrated by parts over x:

$$\begin{aligned} &\int_{y=-\infty}^{y=x} y f(y)^\alpha dF(y) F(x)\Big|_{-\infty}^{\infty} - \int x f(x)^\alpha F(x) dF(x) \\ &= \int y f(y)^\alpha dF(y) - \int x f(x)^\alpha F(x) dF(x) \\ &= \int y f(y)^\alpha\,(1 - F(y))\, dF(y). \end{aligned} \tag{52}$$

The last term in (51) can also be integrated by parts over x as follows:

$$
\begin{aligned}
\int\limits_x \int_{y=x}^{\infty} x f(y)^{\alpha} dF(y) dF(x) &= \int\limits_x \int_{y=x}^{\infty} f(y)^{\alpha} dF(y)\, x\, dF(x) \\
&= \mu^*(x) \int_{y=x}^{\infty} f(y)^{\alpha} dF(y) \Big|_{x=-\infty}^{x=\infty} + \int\limits_x \mu^*(x) f(x)^{\alpha} dF(x) \\
&= \int\limits_y \mu^*(y) f(y)^{\alpha} dF(y), \qquad (53)
\end{aligned}
$$

where $\mu^*(x) = \int_{-\infty}^{x} z dF(z)$ is a partial mean. Adding terms yields (49), and completes the proof. ▮

Suppose that we estimate $P_\alpha(F)$ using a random sample of n iid observations of income y_i, $i = 1, ..., n$, drawn from the distribution $F(y)$ and ordered such that $y_1 \leq y_2 \leq ... \leq y_n$. As in (9), a natural estimator of $P_\alpha(F)$ is $P_\alpha(\hat{F})$, given by substituting the distribution function $F(y)$ by the empirical distribution function $\widehat{F}(y)$, by replacing $f(y)^\alpha$ by a suitable estimator $\widehat{f}(y)^\alpha$ (to be examined below), and by replacing $a(y)$ by $\widehat{a}(y)$:

$$
(54) \quad P_\alpha(\widehat{F}) = \int \widehat{f}(y)^{\alpha} \widehat{a}(y) d\widehat{F}(y) = n^{-1} \sum_{i=1}^{n} \widehat{f}(y_i)^{\alpha} \widehat{a}(y_i),
$$

with $\widehat{p}_\alpha(y_i) = \widehat{f}(y_i)^\alpha \widehat{a}(y_i)$. Note that y_i is the empirical quantile for percentiles between $(i-1)/n$ and i/n. Hence, we may use

$$
(55) \quad \widehat{F}(y_i) = \frac{1}{2}\left(\frac{(i-1)}{n} + \frac{(i)}{n}\right) = 0.5 n^{-1} (2i - 1)
$$

and

$$
(56) \quad \widehat{\mu}^*(y_i) = n^{-1} \left(\sum_{j=1}^{i-1} y_j + \frac{i-(i-1)}{2} y_i \right),
$$

and thus define $\widehat{a}(y_i)$ as

$$
(57) \quad \widehat{a}(y_i) = \widehat{\mu} + y_i \left(n^{-1}(2i-1) - 1\right) - n^{-1} \left(2 \sum_{j=1}^{i-1} y_j + y_i \right),
$$

where $\widehat{\mu}$ is the sample mean.

Observe that adding an exact replication of the sample to the original sample should not change the value of the estimator $P_\alpha(\widehat{F})$. Indeed, supposing that the estimators

$\widehat{f}(\cdot)^{\alpha}$ are invariant to sample size, this is indeed the case when formulae (9) and (57) are used. We record this formally as

OBSERVATION **2.** *Let* $\mathbf{y} = (y_1, y_2, ..., y_n)$ *and* $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_{2n})$ *be two vectors of sizes* n *and* $2n$ *respectively, ordered along increasing values of income. Suppose that for each* $i \in \{1, \ldots, n\}$, $y_i = \tilde{y}_{2i-1} = \tilde{y}_{2i}$ *for all* $i = 1, ..., n$. *Let* $P_\alpha(F_{\mathbf{y}})$ *be the polarization index defined by (9) and (57) for a vector of income* $\mathbf{y}$. *Then, provided that* $f_{\mathbf{y}}(y_i) = f_{\tilde{\mathbf{y}}}(y_i)$ *for* $i = 1, \ldots, n$, *it must be that* $P_\alpha(F_{\mathbf{y}}) = P_\alpha(F_{\tilde{\mathbf{y}}})$.

Proof of Observation 2. It will be enough to show that $2a_{\mathbf{y}}(y_i) = a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1}) + a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i})$ since we have assumed that $f_{\mathbf{y}}(y_i) = f_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1}) = f_{\tilde{\mathbf{y}}}(\tilde{y}_{2i})$ for all $i = 1, \ldots, n$. Clearly, $\mu_{\mathbf{y}} = \mu_{\tilde{\mathbf{y}}}$. Note also that $a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1})$ can be expressed as

$$(58)\quad a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1}) = \mu_{\mathbf{y}} + y_i \left((2n)^{-1} (2(2i-1)-1) - 1\right) - (2n)^{-1} \left(2\sum_{j=1}^{2i-2} \tilde{y}_j + \tilde{y}_{2i-1}\right).$$

Similarly, for $a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i})$, we have

$$(59)\quad a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i}) = \mu_{\mathbf{y}} + y_i \left((2n)^{-1} (2(2i)-1) - 1\right) - (2n)^{-1} \left(2\sum_{j=1}^{2i-1} \tilde{y}_j + \tilde{y}_{2i}\right).$$

Summing (58) and (59), we find

$$\begin{aligned} a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1}) + a_{\tilde{\mathbf{y}}}(\tilde{y}_{2i}) &= 2\left(\mu_{\mathbf{y}} + y_i\left(n^{-1}(2i-1)-1\right) - n^{-1}\left(2\sum_{j=1}^{i-1} y_j + y_i\right)\right) \\ &= 2a_{\mathbf{y}}(y_i). \end{aligned} \quad (60)$$

Adding up the product of $f_{\mathbf{y}}(\tilde{y}_j)a_{\tilde{\mathbf{y}}}(\tilde{y}_j)$ across j and dividing by $2n$ shows that $P_\alpha(F_{\mathbf{y}}) = P_\alpha(F_{\tilde{\mathbf{y}}})$. ∎

When observations are weighted (or "grouped"), with w_i being the sampling weight on observation i and with $\overline{w} = \sum_{j=1}^{n} w_j$ being the sum of weights, a population-invariant definition of $\widehat{g}(y_i)$ is then:

$$(61)\quad \widehat{a}(y_i) = \widehat{\mu} + y_i \left(\overline{w}^{-1}\left(2\sum_{j=1}^{i} w_j - w_i\right) - 1\right) - \overline{w}^{-1}\left(2\sum_{j=1}^{i-1} w_j y_j + w_i y_i\right).$$

(57) is a special case of (61) obtained when $w_i = 1$ for all i. For analytical simplicity, we focus on the case of samples with unweighted iid observations.

$f(y_i)^\alpha$ and the Sampling Distribution of $P_\alpha(\widehat{F})$

It will be generally desirable to adjust our estimator of $f(y_i)^\alpha$ to sample size in order to minimize the sampling error of estimating the polarization indices. To facilitate a more detailed discussion of this issue, first decompose the estimator $P_\alpha(\widehat{F})$ across its separate sources of sampling variability:

$$\begin{aligned} P_\alpha(\widehat{F}) - P_\alpha(F) = &\int (\widehat{p}_\alpha(y) - p_\alpha(y))\, dF(y) + \int p_\alpha(y) d(\widehat{F} - F)(y) \\ &+ \int (\widehat{p}_\alpha(y) - p_\alpha(y))\, d(\widehat{F} - F)(y). \end{aligned} \tag{62}$$

The first source of variation, $\widehat{p}_\alpha(y) - p_\alpha(y)$, comes from the sampling error made in estimating the identification and the alienation effects at each point y in the income distribution. It can be decomposed further as:

$$\begin{aligned} \widehat{p}_\alpha(y) - p_\alpha(y) = &\left(\widehat{f}(y)^\alpha - f(y)^\alpha\right) a(y) + f(y)^\alpha \left(\widehat{a}(y) - a(y)\right) \\ &+ \left(\widehat{f}(y)^\alpha - f(y)^\alpha\right) \left(\widehat{a}(y) - a(y)\right). \end{aligned} \tag{63}$$

As can be seen by inspection, $\widehat{a}(y) - a(y)$ is of order $O(n^{-1/2})$. Assuming that $\widehat{f}(y)^\alpha - f(y)^\alpha$ vanishes as n tends to infinity (as will be shown in the proof of Theorem 2), the last term in (63) is of lower order than the others and can therefore be ignored asymptotically.

This argument also shows that $\widehat{p}_\alpha(y) - p_\alpha(y) \sim o(1)$. Because $F(y) - \widehat{F}(y) = O(n^{-1/2})$, the last term in (62) is of order $o(n^{-1/2})$ and can also be ignored. Combining (62) and (63), we thus see that for large n,

$$P_\alpha(\widehat{F}) - P_\alpha(F) \cong \int \left(\widehat{f}(y)^\alpha - f(y)^\alpha\right) a(y) dF(y) \tag{64}$$

$$+ \int f(y)^\alpha \left(\widehat{a}(y) - a(y)\right) dF(y) \tag{65}$$

$$+ \int p_\alpha(y) d(\widehat{F} - F)(y). \tag{66}$$

The terms (65) and (66) are further developed in the proof of Theorem 2 below.

We thus turn to the estimation of $f(y)^\alpha$ in (64), which we do nonparametrically using kernel density estimation. This uses a kernel function $K(u)$, defined such that $\int_{-\infty}^{\infty} K(u) du = 1$ (this guarantees the desired property that $\int_{-\infty}^{\infty} \widehat{f}(y) dy = 1$) and $K(u) \geq 0$ (this guarantees that $\widehat{f}(y) \geq 0$). It is also convenient to choose a kernel function that is symmetric around 0, with $\int uK(u)du = 0$ and $\int u^2 K(u)du = \sigma_K^2 < \infty$. The estimator $\widehat{f}(y)$ is then defined as

$$(67)\quad \widehat{f}(y) \equiv n^{-1} \sum_{i=1}^{n} K_h\left(y - y_i\right),$$

where $K_h(z) \equiv h^{-1} K(z/h)$. The parameter h is usually referred to as the bandwidth (or window width, or smoothing parameter). One kernel function that has nice continuity and differentiability properties is the Gaussian kernel, defined by

$$(68)\quad K(u) = (2\pi)^{-0.5} \exp^{-0.5u^2},$$

a form that we have used in the illustration.[19]

With $f(y)^\alpha$ estimated according to this general technique, we have the following theorem on the asymptotic sampling distribution of $\widehat{P}_\alpha$.

THEOREM **2.** *Assume that the order-2 population moments of y, $p_\alpha(y)$, $f(y)^\alpha$, $\int_{-\infty}^{y} z f(z)^\alpha dF(z)$ and $y \int_{-\infty}^{y} f(z)^\alpha dF(z)$ are finite. Let h in $K_h(\cdot)$ vanish as n tends to infinity. Then $n^{0.5}\left(P_\alpha(\widehat{F}) - P_\alpha(F)\right)$ has a limiting normal distribution $N(0, V_\alpha)$, with*

$$(69)\quad V_\alpha = \underset{f(y)}{var}\left(v_\alpha(y)\right),$$

where

$$(70)\quad v_\alpha(y) = (1+\alpha)p_\alpha(y) + y \int f(x)^\alpha dF(x) + 2\int_y^\infty (x-y)\, f(x)^\alpha dF(x).$$

Proof of Theorem 2. Consider first (64). Note that

$$\begin{aligned}
\int \left(\widehat{f}(y)^\alpha - f(y)^\alpha\right) a(y)dF(y) &\cong \int \alpha f(y)^{\alpha-1}\left(\widehat{f}(y) - f(y)\right) a(y)dF(y) \\
&= \alpha \int p_{\alpha-1}(y) n^{-1} \sum_{i=1}^{n} K_h\left(y - y_i\right) dF(y) - \alpha \int p_\alpha(y) dF(y) \\
&= \alpha n^{-1} \sum_{i=1}^{n} \int p_{\alpha-1}(y) K_h\left(y - y_i\right) dF(y) - \alpha \int p_\alpha(y) dF(y). \qquad (71)
\end{aligned}$$

Taking $h \to 0$ as $n \to \infty$, and recalling that $\int K_h\left(y - y_i\right) dy = 1$, the first term in (71) tends asymptotically to

$$\alpha n^{-1} \sum_{i=1}^{n} \int p_{\alpha-1}(y) K_h\left(y - y_i\right) dF(y) \cong \alpha n^{-1} \sum_{i=1}^{n} p_{\alpha-1}(y_i) f(y_i) = \alpha n^{-1} \sum_{i=1}^{n} p_\alpha(y_i).$$

[19]Note that the Gaussian kernel has the property that $\sigma_K^2 = 1$.

Thus, we can rewrite the term on the right-hand side of (64) as

$$\int \left(\widehat{f}(y)^\alpha - f(y)^\alpha\right) a(y)dF(y) \cong \alpha n^{-1} \sum_{i=1}^{n} (p_\alpha(y_i) - P_\alpha) = O(n^{-1/2}).$$

Now turn to (65). Let I be an indicator function that equals 1 if its argument is true and 0 otherwise. We find:

$$\begin{aligned}
&\int f(y)^\alpha \left(\widehat{a}(y) - a(y)\right) dF(y) \\
=&\int f(y)^\alpha \left[\left(\widehat{\mu} + y\left(2\widehat{F}(y) - 1\right) - 2\widehat{\mu}^*(y)\right) - a(y)\right] dF(y) \\
\cong&\int f(y)^\alpha \left(n^{-1} \textstyle\sum_{i=1}^{n} \{y_i + y\,(2I[y_i \le y] - 1) - 2y_i I[y_i \le y]\} - a(y)\right) dF(y) \\
=&\, n^{-1} \textstyle\sum_{i=1}^{n} \displaystyle\int f(y)^\alpha \left(y_i \left[1 - 2I[y_i \le y]\right] + 2yI[y_i \le y]\right) dF(y) \\
&\quad - \int f(y)^\alpha \left(\mu + 2yF(y) - 2\mu^*(y)\right) dF(y) \\
=&\, n^{-1} \textstyle\sum_{i=1}^{n} \displaystyle\left(\int f(y)^\alpha dF(y)\, y_i - 2y_i \int_{y_i}^{\infty} f(y)^\alpha dF(y) + 2\int_{y_i}^{\infty} y f(y)^\alpha dF(y)\right) \\
&- \int f(y)^\alpha \left(\mu + 2yF(y) - 2\mu^*(y)\right) dF(y) \\
=&\, O(n^{-1/2}).
\end{aligned}$$

Now consider (66):

$$\int p_\alpha(y) d\left(\widehat{F} - F\right)(y) = n^{-1} \textstyle\sum_{i=1}^{n} \left(f(y_i)^\alpha a(y_i) - P_\alpha\right) = O(n^{-1/2}).$$

Collecting and summarizing terms, we obtain:

$$\begin{aligned}
P_\alpha(\widehat{F}) - P_\alpha(f) \cong n^{-1} \textstyle\sum_{i=1}^{n} &\displaystyle\left((1+\alpha) f(y_i)^\alpha a(y_i) + \int y_i f(y)^\alpha dF(y) + 2\int_{y_i}^{\infty} (y - y_i)\, f(y)^\alpha dF(y)\right) \\
&- \left((1+\alpha)P_\alpha(f) + \int f(y)^\alpha \left(\mu + 2(yF(y) - \mu^*(y))\right) dF(y)\right).
\end{aligned}$$

Applying the law of large numbers to $P_\alpha(\widehat{F}) - P_\alpha(f)$, note that $\lim_{n\to\infty} \mathbf{E}\left[n^{0.5}\left(P_\alpha(\widehat{F}) - P_\alpha(f)\right)\right] = 0$. The central limit theorem then leads to the finding that $n^{0.5}\left(P_\alpha(\widehat{F}) - P_\alpha(f)\right)$ has a limiting normal distribution $N(0, V_\alpha)$, with V_α as described in the statement of the theorem. ▮

Observe that the assertion of Theorem 2 is distribution-free since everything in (69) can be estimated consistently *without* having to specify the population distribution from which the sample is drawn. $P_\alpha(\widehat{F})$ is thus a root-n consistent estimator of $P_\alpha(F)$, unlike the usual non-parametric density and regression estimators which are often $n^{2/5}$ consistent. The strength of Theorem 2 also lies in the fact that so long as h tends to vanish as n increases, the precise path taken by h has a negligible influence on the asymptotic variance since it does not appear in (69).

The Minimization of Sampling Error

In finite samples, however, $P_\alpha(\widehat{F})$ is biased. The bias arises from the smoothing techniques employed in the estimation of the density function $f(y)$. In addition, the finite-sample variance of $P_\alpha(\widehat{F})$ is also affected by the smoothing techniques. As is usual in the non-parametric literature, the larger the value of h, the larger the finite-sample bias, but the lower is the finite-sample variance. We exploit this tradeoff to choose an "optimal" bandwidth for the estimation of $P_\alpha(\widehat{F})$, which we denote by $h^*(n)$.

A common technique is to select $h^*(n)$ so as to minimize the mean square error (MSE) of the estimator, given a sample of size n. To see what this entails, decompose (for a given h) the MSE into the sum of the squared bias and of the variance involved in estimating $P_\alpha(F)$:

$$(72)\quad \mathrm{MSE}_h(P_\alpha(\widehat{F})) = \left(\mathrm{bias}_h\left(P_\alpha(\widehat{F})\right)\right)^2 + \mathrm{var}_h\left(P_\alpha(\widehat{F})\right),$$

and denote by $h^*(n)$ the value of h which minimizes $\mathrm{MSE}_h(P_\alpha(\widehat{F}))$. This value is described in the following theorem:

Theorem **3.** *For large n, $h^*(n)$ is given by*

$$(73)\quad h^*(n) = \sqrt{-\frac{\mathrm{cov}\left(v_\alpha(y), p''_\alpha(y)\right)}{\alpha\sigma_K^2\left(\int f''(y)p_\alpha(y)dy\right)^2}}\, n^{-0.5} + O\left(n^{-1}\right).$$

Proof of Theorem 3. Using (64)–(66), we may write $\mathrm{bias}_h(\widehat{F}_\alpha) = \mathrm{E}\left[P_\alpha(\widehat{F}) - P_\alpha(f)\right]$

as:

$$\mathbf{E}\left[P_\alpha(\widehat{F})-P_\alpha(f)\right] \cong \int \mathbf{E}\left[\widehat{f}(y)^\alpha - f(y)^\alpha\right] a(y)dF(y)$$
$$+\int f(y)^\alpha \mathbf{E}\left[\widehat{a}(y)-a(y)\right]dF(y) + \int p_\alpha(y) d\mathbf{E}\left[\widehat{F}-F\right](y)$$
$$(74) \qquad = \int \mathbf{E}\left[\widehat{f}(y)^\alpha - f(y)^\alpha\right] a(y)dF(y),$$

since $\widehat{a}(y)$ and $\widehat{F}(y)$ are unbiased estimators of (y) and $F(y)$ respectively. For $\mathbf{E}\left[\widehat{f}(y)^\alpha - f(y)^\alpha\right]$, we may use a first-order Taylor expansion around $f(y)^\alpha$:

$$\mathbf{E}\left[\widehat{f}(y)^\alpha - f(y)^\alpha\right] \cong \alpha f(y)^{\alpha-1}\mathbf{E}\left[\widehat{f}(y)-f(y)\right].$$

For symmetric kernel functions, the bias $\mathbf{E}\left[\widehat{f}(y)-f(y)\right]$ can be shown to be approximately equal to (see for instance Silverman (1986, p.39))

$$(75) \quad 0.5h^2\sigma_K^2 f''(y),$$

where $f''(y)$ is the second-order derivative of the density function. Hence, the bias $\mathbf{E}\left[P_\alpha(\widehat{F})-P_\alpha(f)\right]$ is approximately equal to

$$(76) \quad \mathbf{E}\left[P_\alpha(\widehat{F})-P_\alpha(f)\right] \cong 0.5\alpha\sigma_K^2 h^2 \int f''(y)p_\alpha(y)dy = O(h^2).$$

It follows that the bias will be low if the kernel function has a low variance σ_K^2: it is precisely then that the observations "closer" to y will count more, and those are also the observations that provide the least biased estimate of the density at y. But the bias also depends on the curvature of $f(y)$, as weighted by $p_\alpha(y)$: in the absence of such a curvature, the density function is linear and the bias provided by using observations on the left of y is just (locally) outweighed by the bias provided by using observations on the right of y. For the variance $\mathrm{var}_h\left(P_\alpha(\widehat{f})\right)$, we first reconsider the first term in (71), which is the dominant term through which the choice of h influences $\mathrm{var}\left(P_\alpha(\widehat{f})\right)$. We may write this as follows:

$$\alpha n^{-1}\sum_{i=1}^{n}\int p_\alpha(y)K_h\left(y-y_i\right)dy = \alpha n^{-1}\sum_{i=1}^{n}\int p_\alpha(y_i-ht)K\left(t\right)dt$$
$$\cong \alpha n^{-1}\sum_{i=1}^{n}\int K\left(t\right)\left(p_\alpha(y_i)-htp'_\alpha(y_i)+0.5h^2t^2p''_\alpha(y_i)\right)dt$$
$$= \alpha n^{-1}\sum_{i=1}^{n}\left(p_\alpha(y_i)+0.5\sigma_K^2h^2p''_\alpha(y_i)\right), \tag{77}$$

where the first equality substitutes t for $h^{-1}(y_i-y)$, where the succeeding approximation is the result of Taylor-expanding $p_\alpha(y_i-ht)$ around $t=0$, and where the last line follows from the properties of the kernel function $K(t)$. Thus, combining (77) and (69) to incorporate a finite-sample correction for the role of h in the variance of $\widehat{f}_\alpha$, we can write:

$$\text{var}_h\left(P_\alpha(\widehat{f})\right) = n^{-1}\underset{f(y)}{\text{var}}\left(0.5\alpha\sigma_K^2h^2p''_\alpha(y)+v_\alpha(y)\right) = O(n^{-1}).$$

For small h, the impact of h on the finite sample variance comes predominantly from the covariance between $v_\alpha(y)$ and $p''_\alpha(y)$ since $\text{var}\left(0.5\alpha\sigma_K^2h^2p''_\alpha(y)\right)$ is then of smaller order h^4. This covariance, however, is not easily unravelled. When the covariance is negative (which we do expect to observe), a larger value of h will tend to decrease $\text{var}_h\left(P_\alpha(\widehat{f})\right)$ since this will tend to level the distribution of $0.5\alpha\sigma_K^2h^2p''_\alpha(y)+v_\alpha(y)$, which is the random variable whose variance determines the sampling variance of $P_\alpha(\widehat{f})$. Combining squared-bias and variance into (72), we obtain:

$$\text{MSE}_h(P_\alpha(\widehat{f})) = \left(0.5\alpha\sigma_K^2h^2\int f''(y)p_\alpha(y)dy\right)^2 + n^{-1}\underset{f(y)}{\text{var}}\left(0.5\alpha\sigma_K^2h^2p''_\alpha(y)+v_\alpha(y)\right).$$

$h^*(n)$ is found by minimizing $\text{MSE}_h(P_\alpha(\widehat{f}))$ with respect to h. The derivative of $\text{MSE}_h(P_\alpha(\widehat{f}))$ with respect to h gives:

$$h^3\left[\alpha\sigma_K^2\int f''(y)p_\alpha(y)dy\right]^2 + n^{-1}\alpha\sigma_K^2h\int\Big[\left(0.5\alpha\sigma_K^2h^2p''_\alpha(y)+v_\alpha(y)\right)$$
$$-\left(0.5\alpha\sigma_K^2h^2\int p''_\alpha(y)dF(y)+\int v_\alpha(y)dF(y)\right)\Big]\left[p''_\alpha(y)-\int p''_\alpha(y)dF(y)\right]dF(y).$$

Since $h^*(n)>0$ in finite samples, we may divide the above expression by h, and then find $h^*(n)$ by setting the result equal to 0. This yields:

$$h^*(n)^2 = -\frac{n^{-1}\text{cov}\left(v_\alpha(y),p''_\alpha(y)\right)}{\alpha\sigma_K^2\left(\left(\int f''(y)p_\alpha(y)dy\right)^2-0.5n^{-1}\text{var}\left(p''(y)p_\alpha(y)\right)\right)} \tag{78}$$

For large n (and thus for a small optimal h), $h^*(n)$ is thus given by

$$(79)\quad h^*(n) = \sqrt{-\frac{\operatorname{cov}\left(v_\alpha(y), p''_\alpha(y)\right)}{\alpha\sigma_K^2\left(\int f''(y)p_\alpha(y)dy\right)^2}n^{-0.5}} + O\left(n^{-1}\right)$$

This completes the proof. ▮

It is well known that $f''(y)$ is proportional to the bias of the estimator $\hat{f}(y)$. A large value of $\alpha\sigma_K^2\left(\int f''(y)p_\alpha(y)dy\right)^2$ will thus necessitate a lower value of $h^*(n)$ in order to reduce the bias. Conversely, a larger negative correlation between $v_\alpha(y)$ and $p''_\alpha(y)$ will militate in favor of a larger $h^*(n)$ in order to decrease the sampling variance. More importantly, the optimal bandwidth for the estimation of the polarization index is of order $O(n^{-1/2})$, unlike the usual kernel estimators which are of significantly larger order $O(n^{-1/5})$. Because of this, we may expect the precise choice of h not to be overly influential on the sampling precision of polarization estimators.

To compute $h^*(n)$, two general approaches can be followed. We can assume that $f(y)$ is not too far from a parametric density function, such as the normal or the log-normal, and use (73) to compute $h^*(n)$ (for instance, in the manner of Silverman (1986, p.45) for point density estimation). Alternatively, we can estimate the terms in (73) directly from the empirical distribution, using an initial value of h to compute the $f(y)$ in the $v_\alpha(y)$ and $p_\alpha(y)$ functions. For both of these approaches (and particularly for the last one), expression (73) is clearly distribution specific, and it will also generally be very cumbersome to estimate.

It would thus seem useful to devise a "rule-of-thumb" formula that can be used to provide a readily-computable value for h. When the true distribution is that of a normal distribution with variance σ^2, and when a Gaussian kernel (see (68)) is used to estimate $\hat{f}(y)$, h^* is approximately given by:

$$(80)\quad h^* \cong 4.7\, n^{-0.5}\, \sigma\alpha^{0.1}.$$

Easily computed, this formula works well with the normal distribution[20] since it is never farther than 5% from the h^* that truly minimizes the MSE. The use of such approximate rules also seems justified by the fact that the MSE of the polarization indices

[20]Extensive numerical simulations were made using various values of $n \geq 500$, σ and $\alpha = 0.25$ to 1.

does not appear to be overly sensitive to the choice of the bandwidth h. (80) seems to perform relatively well with other distributions than the normal, including the popular log-normal one, although this is less true when the distribution becomes very skewed. For skewness larger than about 6, a more robust — though more cumbersome — approximate formula for the computation of h^* is given by

$$(81) \quad h^* \cong n^{-0.5} IQ \frac{(3.76 + 14.7\sigma_{ln})}{(1 + 1.09 \cdot 10^{-4}\sigma_{ln})^{(7268+15323\alpha)}},$$

where IQ is the interquartile and σ_{ln} is the variance of the logarithms of income – an indicator of the skewness of the income distribution.

References

Aczél, J. (1966): *Lectures on Functional Equations and Their Applications*, New York: Academic Press.

Alesina, A. and E. Spolaore (1997) "On the Number and Size of Nations," *Quarterly Journal of Economics* **113**, 1027–1056.

Anderson, G. (1996). "Nonparametric Tests of Stochastic Dominance In Income Distributions," *Econometrica*, **64**, 1183-1193.

Beach, C. and R. Davidson, N. (1983): "Distribution-Free Statistical Inference with Lorenz Curves and Income Shares," *Review of Economic Studies*, **50 (4)**, 723–35.

Beach, C. and J. Richmond (1985): "Joint Confidence Intervals for Income Shares and Lorenz," *International Economic Review*, **26 (2)**, 439–50.

Bishop, J.A., S. Chakraborti and P.D. Thistle (1989): ' 'Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves", *The Review of Economics and Statistics*, **71** , 725–27.

Chakravarty, S.R. and A. Majumder (2001): "Inequality, Polarization and Welfare: Theory and Applications," *Australian Economic Papers* **40**, 1–13.

Collier, P. and A. Hoeffler (2001): "Greed and Grievance in Civil War," mimeo., World Bank.

Davidson, R. and J.Y. Duclos (1997): "Statistical Inference for the Measurement of the Incidence of Taxes and Transfers," *Econometrica*, **65 (6)**, 1453–1465.

——— and ——— (2000): "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 68, 1435–1465.

D'Ambrosio, C. and E. Wolff (2001): "Is Wealth Becoming More Polarized in the United States?" Working Paper 330, Levy Economics Institute, Bard College.

Duclos, J.Y., Esteban, J., and D. Ray (2003): "Polarization: Concepts, Measurement, Estimation," mimeo., `http://www.econ.nyu.edu/user/debraj/papers/DuclosEstebanRay.pdf`.

Esteban, J. and D. Ray (1991) "On the Measurement of Polarization," Boston University, Institute for Economic Development, Working Paper 18.

——— and ——— (1994) "On the Measurement of Polarization," *Econometrica* **62**, 819–852.

——— and ——— (1999) "Conflict and Distribution," *Journal of Economic Theory* **87**, 379–415.

——— and ——— (2001): "Collective Action and the Group Size Paradox," *American Political Science Review* **95**, 663–672.

Esteban, J., C. Gradín and D. Ray (1998): "Extensions of a Measure of Polarization, with an Application to the Income Distribution of Five OECD countries," mimeo., Instituto de Análisis Económico.

Fajnzylber, P. Lederman, D. and N. Loayza (2000): "Crime and Victimization: An Economic Perspective," *Economia* **1**, 219–278.

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

Garcia-Montalvo, J. and M. Reynal-Querol (2002): "Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth," mimeo.

Gradín, C. (2000): "Polarization by Sub-Populations in Spain, 1973–91," *Review of Income and Wealth* **46**, 457–474.

Härdle, W. (1990): *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.

Kakwani, N.C. (1993): "Statistical Inference in the Measurement of Poverty," *Review of Economics and Statistics* **75 (3)**, 632–639.

Knack, S. and P. Keefer (2001): "Polarization, Politics and Property Rights: Links between Inequality and Growth," mimeo., World Bank.

Milanovic, B. (2000): "A New Polarization Measure and Its Applications," mimeo., Development Research Group, World Bank.

Olson, M. (1965): *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Pagan, A. R., and A. Ullah (1999). *Nonparametric Econometrics*, Cambridge: Cambridge University Press.

Pareto, V. (1906): *Manual of Political Economy*, New York: A.M. Kelley, 1927 edition.

Quah, D. (1996): "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics," *Economic Journal* **106**, 1045–1055.

——— (1997): "Empirics for Growth and Distribution: Stratification, Polarization and Convergence Clubs," *Journal of Economic Growth* **2**, 27–59.

Reynal-Querol, M. (2002): "Ethnicity, Political Systems, and Civil Wars," *Journal of Conflict Resolution* **46**, 29–54.

Rodríguez, J.G. and R. Salas (2002): "Extended Bi-Polarization and Inequality Measures," mimeo, Universidad Complutense de Madrid.

Sen, A. (1997): *On Economic Inequality*, Second Edition, Oxford: Oxford University Press and Clarendon Press.

Silverman, B.W. (1986): *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Wang, Y.Q. and Tsui, K.Y. (2000): "Polarization Orderings and New Classes of Polarization Indices," *Journal of Public Economic Theory* **2**, 349–363.

Wolfson, M.C. (1994): "When Inequalities diverge," *American Economic Review* **84**, Papers and Proceedings, 353–358.

——— (1997): "Divergent Inequalities: Theory and Empirical Results," *Review of Income and Wealth* **43**, 401–421.

Zhang, X. and R. Kanbur (2001): "What Difference Do Polarisation Measures Make? An Application to China," *Journal of Development Studies* **37**, 85–98.

TABLE I:
LIS Country Codes

Abbreviations	Countries	Years	Sample Sizes
as	Australia	1989 / 1994	16,331 / 7,441
be	Belgium	1992 / 1997	3,821 / 4,632
cn	Canada	1991 / 1994	21,647 / 40,849
cz	Czech Republic	1992 / 1996	16,234 / 28,148
dk	Denmark	1992 / 1995	12,895 / 13,124
fi	Finland	1991 / 1995	11,749 / 9,263
fr	France	1989 / 1994	9,038 / 11,294
ge	Germany	1989 / 1994	4,187 / 6,045
hu	Hungary	1991 / 1994	2,019 / 1,992
is	Israel	1992 / 1997	5,212 / 5,230
it	Italy	1991 / 1994	8,188 / 8,135
lx	Luxembourg	1991 / 1994	1,957 / 1,813
mx	Mexico	1989 / 1996	11,531 / 14,042
nl	Netherlands	1991 / 1994	4,378 / 5,187
nw	Norway	1991 / 1995	8,073 / 10,127
pl	Poland	1992 / 1995	6,602 / 32,009
rc	Rep. of China / Taiwan	1991 / 1995	16,434 / 14,706
ru	Russia	1992 / 1995	6,361 / 3,518
sw	Sweden	1992 / 1995	12,484 / 16,260
uk	United Kingdom	1991 / 1995	7,056 / 6,797
us	United States	1991 / 1994	16,052 / 66,014

TABLE II:
Polarization Indices and Polarization Rankings (Rkg) from LIS Wave 3

α= Country	Index 0	Rkg	Index 0.25	Rkg	Index 0.50	Rkg	Index 0.75	Rkg	Index 1	Rkg
cz92	0.2082	1	0.1767	1	0.1637	2	0.1585	4	0.1575	11
	0.0023		0.0014		0.0011		0.0011		0.0012	
fi91	0.2086	2	0.1782	2	0.1611	1	0.1505	1	0.1436	1
	0.0017		0.0010		0.0007		0.0005		0.0005	
be92	0.2236	3	0.1898	4	0.1699	4	0.1571	3	0.1484	3
	0.0028		0.0018		0.0012		0.0010		0.0010	
sw92	0.2267	4	0.1888	3	0.1674	3	0.1543	2	0.1459	2
	0.0019		0.0012		0.0008		0.0006		0.0006	
nw91	0.2315	5	0.1919	5	0.1713	5	0.1588	5	0.1505	5
	0.0029		0.0017		0.0013		0.0011		0.0011	
dk92	0.2367	6	0.1964	6	0.1744	6	0.1603	6	0.1504	4
	0.0026		0.0015		0.0011		0.0010		0.0011	
lx91	0.2389	7	0.2002	7	0.1787	8	0.1652	8	0.1563	10
	0.0051		0.0032		0.0024		0.0022		0.0023	
ge89	0.2469	8	0.2019	8	0.1779	7	0.1634	7	0.1540	7
	0.0048		0.0028		0.0021		0.0020		0.0021	
nl91	0.2633	9	0.2122	9	0.1859	9	0.1700	9	0.1596	16
	0.0054		0.0031		0.0024		0.0024		0.0025	
rc91	0.2708	10	0.2189	10	0.1902	11	0.1723	14	0.1603	17
	0.0019		0.0011		0.0009		0.0008		0.0009	
pl92	0.2737	11	0.2193	11	0.1894	10	0.1706	11	0.1577	13
	0.0032		0.0019		0.0014		0.0012		0.0013	
fr89	0.2815	12	0.2229	12	0.1912	12	0.1715	12	0.1580	14
	0.0033		0.0019		0.0014		0.0013		0.0014	
hu91	0.2828	13	0.2230	13	0.1913	13	0.1719	13	0.1587	15
	0.0066		0.0039		0.0028		0.0026		0.0027	
it91	0.2887	14	0.2307	15	0.1968	15	0.1741	15	0.1577	12
	0.0028		0.0016		0.0012		0.0011		0.0012	
cn91	0.2891	15	0.2301	14	0.1945	14	0.1701	10	0.1523	6
	0.0018		0.0011		0.0008		0.0006		0.0006	
is92	0.3055	16	0.2421	17	0.2051	17	0.1804	18	0.1626	18
	0.0036		0.0021		0.0016		0.0015		0.0015	
as89	0.3084	17	0.2421	16	0.2023	16	0.1750	16	0.1549	8
	0.0020		0.0012		0.0008		0.0007		0.0008	
uk91	0.3381	18	0.2607	18	0.2185	19	0.1911	19	0.1716	19
	0.0053		0.0028		0.0023		0.0023		0.0025	
us91	0.3394	19	0.2625	19	0.2140	18	0.1802	17	0.1551	9
	0.0019		0.0012		0.0008		0.0006		0.0006	
ru92	0.4017	20	0.2957	20	0.2400	20	0.2046	20	0.1797	20
	0.0066		0.0035		0.0029		0.0029		0.0031	
mx89	0.4909	21	0.3462	21	0.2802	21	0.2432	21	0.2202	21
	0.0055		0.0034		0.0030		0.0032		0.0036	

Standard errors appear on every second line.

TABLE III:
Polarization Indices and Polarization Rankings (Rkg) from LIS Wave 4

	Index	Rkg	Index	Rkg	Index	Rkg	Index	Rkg	Index	Rkg
$\alpha=$	0		0.25		0.50		0.75		1	
Country										
fi95	0.2174	1	0.1832	1	0.1661	2	0.1564	2	0.1506	6
	0.0027		0.0016		0.0012		0.0011		0.0012	
sw95	0.2218	2	0.1845	2	0.1652	1	0.1549	1	0.1498	3
	0.0019		0.0012		0.0008		0.0007		0.0008	
lx94	0.2353	3	0.1978	4	0.1764	4	0.1633	7	0.1549	8
	0.0043		0.0028		0.0021		0.0017		0.0019	
nw95	0.2403	4	0.1970	3	0.1750	3	0.1616	3	0.1527	7
	0.0049		0.0029		0.0024		0.0023		0.0024	
be97	0.2496	5	0.2061	5	0.1796	5	0.1616	4	0.1486	1
	0.0029		0.0018		0.0012		0.0010		0.0010	
dk95	0.2532	6	0.2073	6	0.1808	6	0.1632	6	0.1504	5
	0.0026		0.0015		0.0011		0.0011		0.0011	
nl94	0.2558	7	0.2094	7	0.1812	7	0.1624	5	0.1491	2
	0.0029		0.0018		0.0012		0.0009		0.0010	
cz96	0.2589	8	0.2104	8	0.1854	9	0.1709	10	0.1618	13
	0.0017		0.0010		0.0008		0.0007		0.0008	
ge94	0.2649	9	0.2133	9	0.1846	8	0.1669	8	0.1553	10
	0.0048		0.0030		0.0023		0.0021		0.0022	
rc95	0.2781	10	0.2234	10	0.1931	10	0.1742	11	0.1614	12
	0.0021		0.0013		0.0009		0.0009		0.0010	
cn94	0.2859	11	0.2289	12	0.1933	11	0.1687	9	0.1504	4
	0.0011		0.0007		0.0005		0.0004		0.0003	
fr94	0.2897	12	0.2284	11	0.1963	12	0.1766	13	0.1634	14
	0.0031		0.0018		0.0014		0.0013		0.0014	
as94	0.3078	13	0.2433	14	0.2033	14	0.1757	12	0.1553	9
	0.0028		0.0016		0.0012		0.0010		0.0011	
pl95	0.3108	14	0.2389	13	0.2023	13	0.1799	14	0.1645	15
	0.0024		0.0014		0.0011		0.0010		0.0011	
hu94	0.3248	15	0.2486	15	0.2087	15	0.1852	15	0.1700	18
	0.0081		0.0048		0.0037		0.0035		0.0038	
is97	0.3371	16	0.2598	17	0.2159	17	0.1871	18	0.1666	17
	0.0044		0.0025		0.0019		0.0018		0.0020	
it95	0.3406	17	0.2596	16	0.2148	16	0.1856	16	0.1647	16
	0.0037		0.0021		0.0016		0.0015		0.0016	
uk95	0.3429	18	0.2622	18	0.2193	18	0.1925	19	0.1741	19
	0.0041		0.0022		0.0018		0.0018		0.0020	
us94	0.3622	19	0.2747	19	0.2223	19	0.1868	17	0.1610	11
	0.0010		0.0006		0.0004		0.0004		0.0004	
ru95	0.4497	20	0.3222	20	0.2566	20	0.2164	20	0.1889	20
	0.0061		0.0035		0.0028		0.0028		0.0030	
mx96	0.4953	21	0.3483	21	0.2826	21	0.2464	21	0.2237	21
	0.0046		0.0028		0.0025		0.0027		0.0030	

Standard errors appear on every second line.

TABLE IV:
Alienation and Identification — LIS Wave 3

$\alpha =$	0	0.25				0.50				0.75				1.00			
Country	Gini	$\bar{\iota}$	c^{a}	$\bar{\iota} \cdot c^{\mathrm{a}}$	P	$\bar{\iota}$	c^{a}	$\bar{\iota} \cdot c^{\mathrm{a}}$	P	$\bar{\iota}$	c^{a}	$\bar{\iota} \cdot c^{\mathrm{a}}$	P	$\bar{\iota}$	c^{a}	$\bar{\iota} \cdot c^{\mathrm{a}}$	P
as89	0.3084	0.8508	0.9227	0.7851	0.2421	0.7440	0.8815	0.6559	0.2023	0.6627	0.8562	0.5675	0.1750	0.5984	0.8394	0.5023	0.1549
be92	0.2233	0.9110	0.9327	0.8497	0.1897	0.8518	0.8931	0.7608	0.1699	0.8105	0.8678	0.7034	0.1571	0.7811	0.8506	0.6643	0.1484
cn91	0.2891	0.8634	0.9219	0.7960	0.2301	0.7658	0.8784	0.6727	0.1945	0.6916	0.8509	0.5885	0.1701	0.6332	0.8321	0.5269	0.1523
cz92	0.2081	0.9504	0.8935	0.8492	0.1767	0.9337	0.8423	0.7865	0.1637	0.9364	0.8132	0.7615	0.1585	0.9526	0.7944	0.7567	0.1575
dk92	0.2367	0.9051	0.9169	0.8298	0.1964	0.8415	0.8759	0.7370	0.1744	0.7952	0.8519	0.6774	0.1603	0.7598	0.8361	0.6352	0.1504
fi91	0.2086	0.9227	0.9259	0.8543	0.1782	0.8747	0.8829	0.7723	0.1611	0.8440	0.8547	0.7214	0.1505	0.8248	0.8345	0.6882	0.1435
fr89	0.2815	0.8782	0.9015	0.7917	0.2229	0.7978	0.8514	0.6792	0.1912	0.7406	0.8224	0.6091	0.1715	0.6979	0.8041	0.5612	0.1580
ge89	0.2469	0.9021	0.9066	0.8179	0.2019	0.8398	0.8583	0.7208	0.1779	0.7984	0.8290	0.6618	0.1634	0.7707	0.8094	0.6238	0.1540
hu91	0.2828	0.8797	0.8965	0.7887	0.2230	0.8007	0.8451	0.6767	0.1913	0.7451	0.8157	0.6078	0.1719	0.7042	0.7972	0.5614	0.1587
is92	0.3055	0.8626	0.9188	0.7926	0.2421	0.7663	0.8761	0.6714	0.2051	0.6944	0.8505	0.5906	0.1804	0.6384	0.8337	0.5322	0.1626
it91	0.2887	0.8676	0.9212	0.7993	0.2307	0.7745	0.8802	0.6817	0.1968	0.7046	0.8558	0.6030	0.1741	0.6501	0.8404	0.5463	0.1577
lx91	0.2389	0.9088	0.9222	0.8381	0.2002	0.8490	0.8807	0.7477	0.1787	0.8081	0.8557	0.6915	0.1652	0.7798	0.8392	0.6544	0.1563
mx89	0.4909	0.8343	0.8453	0.7052	0.3462	0.7302	0.7817	0.5707	0.2802	0.6588	0.7520	0.4954	0.2432	0.6090	0.7366	0.4486	0.2202
nl91	0.2633	0.8952	0.9003	0.8059	0.2122	0.8280	0.8526	0.7059	0.1859	0.7822	0.8255	0.6457	0.1700	0.7499	0.8084	0.6062	0.1596
nw91	0.2315	0.9128	0.9082	0.8290	0.1919	0.8581	0.8623	0.7400	0.1713	0.8216	0.8347	0.6859	0.1588	0.7970	0.8158	0.6502	0.1505
pl92	0.2737	0.8837	0.9067	0.8013	0.2193	0.8068	0.8575	0.6919	0.1894	0.7526	0.8278	0.6230	0.1705	0.7129	0.8081	0.5762	0.1577
rc91	0.2708	0.8883	0.9099	0.8083	0.2189	0.8152	0.8616	0.7024	0.1902	0.7645	0.8323	0.6362	0.1723	0.7281	0.8130	0.5919	0.1603
ru92	0.4017	0.8300	0.8868	0.7361	0.2957	0.7138	0.8369	0.5974	0.2400	0.6282	0.8108	0.5094	0.2046	0.5622	0.7960	0.4475	0.1797
sw92	0.2267	0.9077	0.9177	0.8330	0.1888	0.8499	0.8691	0.7387	0.1674	0.8126	0.8376	0.6807	0.1543	0.7889	0.8159	0.6436	0.1459
uk91	0.3381	0.8521	0.9047	0.7709	0.2607	0.7498	0.8618	0.6461	0.2185	0.6737	0.8390	0.5652	0.1911	0.6145	0.8258	0.5074	0.1716
us91	0.3394	0.8298	0.9320	0.7734	0.2625	0.7063	0.8930	0.6307	0.2140	0.6116	0.8685	0.5311	0.1803	0.5364	0.8520	0.4571	0.1551