

Duclos, Jean-Yves; Esteban, Joan; Ray, Debraj

Working Paper

Polarization: Concepts, measurement, estimation

LIS Working Paper Series, No. 335

Provided in Cooperation with:

Luxembourg Income Study (LIS)

Suggested Citation: Duclos, Jean-Yves; Esteban, Joan; Ray, Debraj (2002) : Polarization: Concepts, measurement, estimation, LIS Working Paper Series, No. 335, Luxembourg Income Study (LIS), Luxembourg

This Version is available at:

<https://hdl.handle.net/10419/95427>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Luxembourg Income Study Working Paper Series

Working Paper No. 335

Polarization: Concepts, Measurement, Estimation

Jean Yves Duclos, Joan Esteban, Debraj Ray

November 2002



Luxembourg Income Study (LIS), asbl

Polarization: Concepts, Measurement, Estimation

Jean-Yves Duclos

CIRPÉE and Université Laval

Joan Esteban

Instituto de Análisis Económico (CSIC) and Universitat Pompeu Fabra

Debraj Ray

New York University and Instituto de Análisis Económico (CSIC)

November 2002

Abstract

The purpose of this paper is two-fold. First, we develop the measurement theory of polarization for the case in which asset distributions can be described using density functions. 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 derived 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.

This research was partly funded by The Pew Charitable Trusts, CRSH and FQRSC. Duclos and Ray thank the Instituto de Análisis Económico (CSIC) for hospitality during the startup phase of this project. Esteban thanks the support from the Instituto de Estudios Fiscales and the MCYT. We are grateful to Nicolas Beaulieu for his excellent research assistance.

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 polarization¹ 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, and that — at least in principle — it could be connected with several aspects of social, economic and political change.²

Following Esteban and Ray (1991, 1994), we rely almost exclusively on what might be called the *identification-alienation* framework. The idea is simple: social 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. Rather, 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 is, at best, orthogonal to the idea 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 asset 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 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), Howes (1993), Kakwani (1993), Anderson

¹See 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).

²See, 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.

(1996), and Davidson and Duclos (1997, 2000)). A rigorous axiomatic development of the polarization concept in the “density case” is a prerequisite for proper statistical examination of polarization.

In this paper we concentrate mainly on the axiomatics and estimation of “pure income polarization”, that is, of indices of polarization for which individuals identify themselves only with those with exactly the same income level. This brings us to the second, predominantly statistical, issue of how the estimation of polarization is to be conducted. The main statistical problem is how to estimate the size of the groups to which individuals belong. Fixing 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, we show that 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 thereafter. 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.³ The practical difficulty is that the population is assumed to have *already* been bunched in the relevant groups. This feature rendered the measure of little use for many interesting problems.⁴ As mentioned above, the present paper solves both problems and provides what we hope is a useable measure. In addition, the main axioms that we use to characterize a measure of income polarization are completely independent from the ones used in ER. We thus find it remarkable 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 present a general introduction to the identity-alienation framework. We consider different notions of identification and alienation, each leading to a different class of measures. One of these cases is that of “pure” income polarization, in which income (or wealth) alone is used to define groups. [In general, other – possibly non-economic – criteria may be used to construct groups.] In Section 3 we axiomatically characterize a measure of pure income polarization. In Section 4, we turn to estimation and inference issues for polarization measures. In Section 5, we illustrate the axiomatic and statistical results using data drawn from the Luxembourg Income Study (LIS) data sets for 21 countries. We compute

³ER (Section 4, p. 846) mention this problem.

⁴In 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.

the Gini coefficient and the polarization measure for these countries for years in Wave 3 (1989–1992) and Wave 4 (1994–1997), and show that the two indices furnish distinct information on the shape of the distributions. All proofs are in Section 6.

2. Identification and Alienation

Although this paper is primarily concerned with income polarization, we begin with a more general discussion. We take a broader perspective in which individual characteristics other than income may matter, and introduce the identification-alienation (IA) framework that informs all of our analysis.

Suppose that society is divided into G groups, the choice of which is left to the applied researcher. These may be income or wealth groups, or divisions based on other criteria: ethnicity, race, religion, caste, gender, region, and so on. The reason why this choice must be left to the researcher is that no amount of measurement theory can tell us *which* classifications are salient for a particular society.

With overall population normalized to unity, let F denote the cdf for overall income distribution in society, and denote by F_j the *unnormalized* distributions for each group j , so that

$$F(x) = \sum_{j=1}^G F_j(x)$$

for each income level x . We assume that this “true” distribution admits a density, which we denote by f for the society as a whole, and f_j (once again unnormalized) for each group j . Let n_j be the population mass and μ_j be the mean income of group j . We will let \mathbf{F} stand for the joint distribution summarized by $(F; F_1, \dots, F_G)$.

2.1. IA and Polarization. Here is the hypothesis underlying all that we do. Each individual is assumed to be subject to two forces: he feels *identification* with those he considers to be members of his “own group”, and alienation from those he considers to be members of “other groups”. Thus, keeping matters deliberately abstract for the moment, we suppose that a member of group j with income x feels *identification* $i = \lambda_j(x, \mathbf{F})$ with individuals in his own group j . We also suppose that he experiences alienation from individuals in other groups (or perhaps the same group): $a = \delta_{jk}(x, y)$ with respect to some member of group k with income y (where δ_{jk} is some group-pair-specific distance function). The point is that our individual’s *effective antagonism* towards the person with income y in group k is his alienation weighted by the identification he feels. That is, alienation increases effective antagonism, but in itself is not enough: for this to translate into social tension, this individual must find like-minded compatriots. At any rate, this is the provisional assumption on which everything here is based (and it is ultimately an empirical question whether the resulting measure has any explanatory value).

Introducing, then, an increasing function $T(i, a)$ to capture effective antagonism, we define polarization to be the “sum” of all effective antagonisms:

$$(1) \quad P(\mathbf{F}) \equiv \sum_j \sum_k \int_x \int_y T(\lambda_j(x, \mathbf{F}), \delta_{jk}(x, y)) dF_j(x) dF_k(y)$$

Described in this way, the measure is not very operational. Much is obviously left to the choice of identification and alienation functions. Part of the goal of thinking axiomatically is to narrow down these choices in some “reasonable” way.

The analysis in Esteban and Ray [1994] (ER) — and in the text that follows — imposes a certain structure on the identification and alienation functions for the special case in which both identification and alienation are based on the same characteristic. This characteristic can be income or wealth as well as any other socio-political feature. The key restriction, however, is that whatever we choose the salient characteristic to be, inter-group alienation has to be fuelled by the very same characteristic. This seems natural 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 does not seem appropriate here to base inter-ethnic alienation as depending on some suitably defined ethnicity distance. In the cases of socially based group identification we consider more natural to take a multi-dimensional approach to polarization, permitting alienation to depend on characteristics other than the one that defines group identity. Because of these arguments we consider that the most relevant instance of unidimensional polarization is the case of economic polarization, whenever this is the appropriate salient feature. We shall thus refer to this case as (pure) *income polarization*. In a later section, we provide axiomatically derived measures for this concept. In this section, however, we liberally transplant our findings to the case of social polarization ($G \geq 2$), but with no further axiomatic reasoning.

2.2. Leading Special Cases.

2.2.1. Income Polarization. First suppose that $G = 1$, so that income is the only variable of interest. What, then, is the source of identification? For any individual, this source must be found in the measure of individuals in his immediate neighborhood, with whom he shares a commonality of income. For a person with income x , let us proxy this by the *density* $f(x)$ at x .⁵ For two individuals x and y , let $\delta(x, y)$ simply stand for the absolute difference $|x - y|$. The polarization measure in (1) then acquires the format

$$(2) \quad P(\mathbf{F}) \equiv \int_x \int_y T(f(x), |x - y|) dF(x) dF(y).$$

Apart from the fact that we are looking at income polarization ($G = 1$), this is not much of a specialization. The function T continues to incorporate a good deal of arbitrariness. What we will show in Theorem 1, however, is that under some axioms on the polarization ordering, the measure in (2) must reduce to

$$(3) \quad P(\mathbf{F}) \simeq \int_x \int_y f(x)^\alpha |x - y| dF(x) dF(y).$$

for some $\alpha > 0$ (in fact, we also deduce some restrictions on α , on which more later).

Notice that if α were to be set equal to zero in this equation, our measure would simply reduce to the Gini coefficient, a well-known inequality measure. To be sure, α *cannot* be set equal to zero: our axioms will rule this out. The extra weight on density summarized by α is *precisely* how identification enters the picture: a larger value of α corresponds to a higher weighting of alienation by identification.

⁵ER, in their discussion of the continuous case, posit a weighting function around the going income which is used to aggregate nearby individuals for purposes of identification. For an appropriately chosen weighting function, our formulation corresponds to the special case in which this weighting function has vanishingly small support.

2.2.2. *Pure Social Polarization.* The form of the polarization measure suggested by (3) can be transplanted to social polarization without difficulty (though, as already remarked, the *axiomatization* of such transplants remains a nontrivial endeavor). Consider, for instance, the case of “pure social polarization”, in which income plays no role. That is, there are G groups. Each person is 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 (formally, $\delta_{jk}(x, y) = \Delta_{jk}$ for every pair of groups j and k no matter what the incomes x and y are). For the special case in which inter-group social distances satisfy additivity⁶ pure social polarization can be captured by the ER measure

$$(4) \quad P_s(\mathbf{F}) = \sum_{j=1}^G \sum_{k=1}^G n_j^\alpha n_k \Delta_{jk}.$$

However, the additivity property of inter-group alienation is not always reasonable. There are 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]), in which one simply sets

$$(5) \quad \Delta_{jk} = 1 \text{ if } j \neq k, \text{ and } = 0 \text{ if } j = k.$$

This yields the following variant of (3), [or equivalently of 4] which we might call *pure social polarization* (P_s):⁷

$$(6) \quad \tilde{P}_s(\mathbf{F}) = \sum_{j=1}^G \sum_{k \neq j} n_j^\alpha n_k.$$

2.2.3. *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. [For instance, both low-income and high-income Hindus may feel antagonistic towards Muslims as a whole while sharing very little in common with each other.] Then we get what one might call *social polarization with income-mediated identification*:

$$(7) \quad P_s(\mathbf{F}) = \sum_{j=1}^G \sum_{k \neq j} \int_x f_j(x)^\alpha n_k dF_j(x).$$

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 demarcating role — they are necessary (but not sufficient) for

⁶That is, if Δ_{ki} and Δ_{ij} are the distance between groups k and i and i and j , respectively, then $\Delta_{kj} = \Delta_{ki} + \Delta_{ij}$.

⁷See 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.

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

$$(8) \quad P^*(\mathbf{F}) = \sum_{j=1}^G \sum_{k \neq j} \int_x \int_y f_j(x)^\alpha |x - y| dF_j(x) dF_k(y).$$

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. In particular, these measures should certainly be used in place of more commonly-used fragmentation measures in the analysis of conflict.

3. Measuring Income Polarization

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

3.1. 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. An individual located at income x is presumed to feel a sense of identification that depends on the density at x , $f(x)$. More generally, one might consider that individuals have a "window of identification". However, the foundations for the width of such identification window seem unclear. We have thus opted for defining our family of polarization measures for the limit case when the window width becomes zero. But, as discussed in Section 3.4, even this seemingly narrow specification has broader implications.

An individual located at x feels alienation $|x - y|$ as far as an individual located at y is concerned. As in ER, we write the *effective antagonism* of x towards y (under f) 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 $T(0, a) = T(i, 0) = 0$, just as in ER. Finally, we take polarization to be proportional to the "sum" of all effective antagonisms:

$$(9) \quad P(F) = \int \int T(f(x), |x - y|) f(x) f(y) dx dy,$$

The idea is to place some axioms on this starting point so as to narrow down the functional form of T .

3.2. Axioms.

3.2.1. Densities and Basic Operations. Our axioms will largely be based on domains that are unions of one or more symmetric "basic densities." The densities will be scaled down and up to accommodate varying populations. The populations that inhabit each basic density will not be normalized in any way. The building block for these densities we will call *kernels*. These are symmetric, unimodal density functions f with compact support that we always situate on the interval $[0, 2]$, so that their mean is one. By symmetry we mean that $f(x) = f(2 - x)$ for all $x \in [0, 1]$, and by unimodality we mean that f is nondecreasing on $[0, 1]$. We take the overall population of a kernel to be unity. To be sure, a kernel f can be *population scaled* to any population p by simply multiplying f pointwise by p to arrive at a new density pf . Likewise, any kernel (or density for that matter) 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 a kernel f can be *income scaled* to any new mean μ that we please as follows: $g(x) = (1/\mu)f(x/\mu)$ for all x . This will give rise to a new density g that has support $[0, 2\mu]$ and a mean of μ .⁸ Any scaling or slide (or combinations thereof) of a kernel we will call a *basic density*. The operations described above are all standard and do not require much explanation. What we shall also use is the notion of a *squeeze*, defined as follows. Let f be any density with mean μ and let λ lie in $(0, 1]$. A λ -squeeze of the density f is a transformation of this density as follows:

$$(10) \quad f^\lambda(x) \equiv \frac{1}{\lambda} f\left(\frac{x - [1 - \lambda]\mu}{\lambda}\right).$$

A (λ) -squeeze is, in words, a very special sort of mean-preserving reduction in the spread of a density. 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.

[P.1] For each $\lambda \in (0, 1)$, f^λ is a density.

[P.2] For each $\lambda \in (0, 1)$, f^λ has the same mean as f .

[P.3] If $0 < \lambda < \lambda' < 1$, then f^λ second-order stochastically dominates $f^{\lambda'}$.

[P.4] 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.

3.2.2. *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 basic 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, 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.

Our next axiom considers an initial situation composed of three disjoint densities, derived from identical kernels as shown in the diagram. The situation is completely symmetric, with densities 1 and 3 having the same total population

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

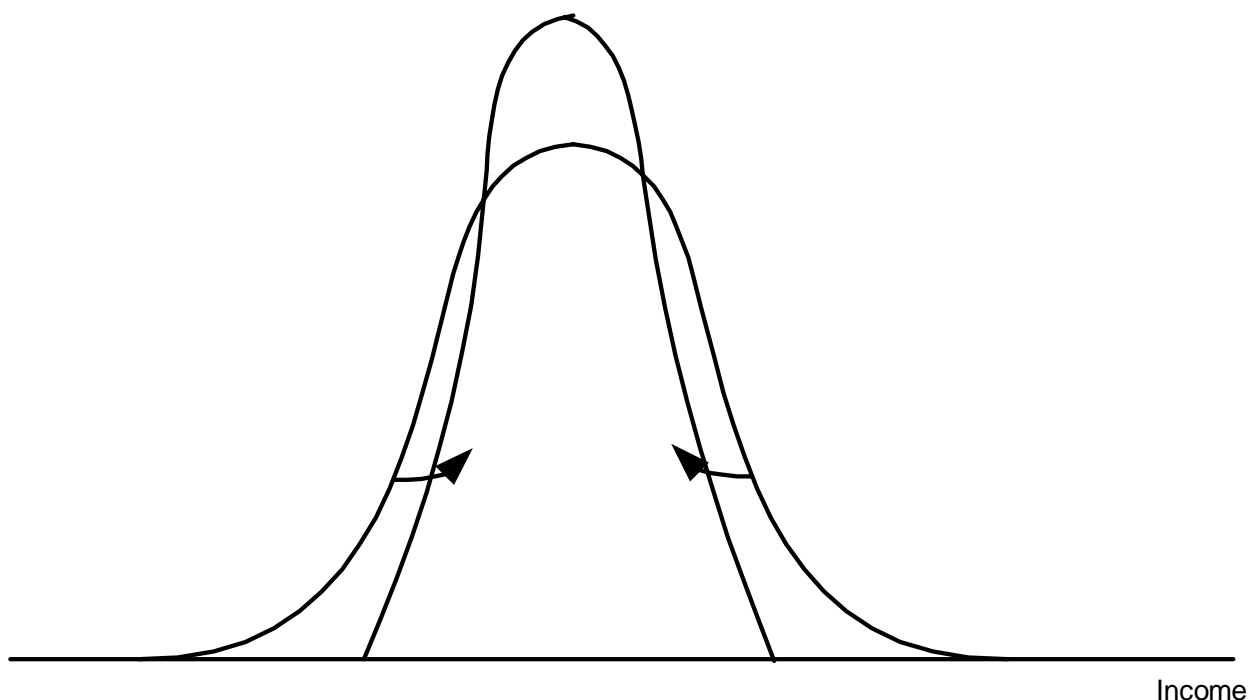


Figure 1: A Single Squeeze Cannot Increase Polarization.

and with density 2 exactly midway between densities 1 and 3. Finally, assume that all supports are disjoint.

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

In some sense, this is the defining axiom of polarization. This is precisely what we 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 generated by the same kernel.

Axiom 3. Consider a symmetric distribution composed of four basic densities drawn from the same kernel, with 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.

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

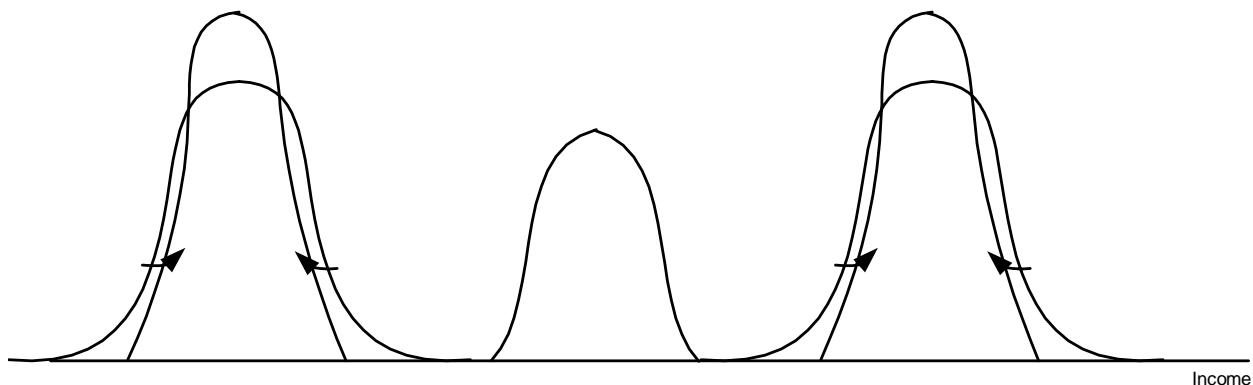


Figure 2: A Double Squeeze Cannot Lower Polarization.

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.

3.3. Characterization Theorem.

Theorem 1. A measure P , as described in (9), satisfies Axioms 1–4 if and only if it is proportional to

$$(11) \quad P_\alpha(F) \equiv \int \int f(x)^{1+\alpha} f(y) |y - x| dy dx,$$

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

3.4. Discussion . 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 by multiplying $P_\alpha(F)$ by $\mu^{\alpha-1}$, where μ is mean income.

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. In fact, we will see in later discussion that there are several other measures

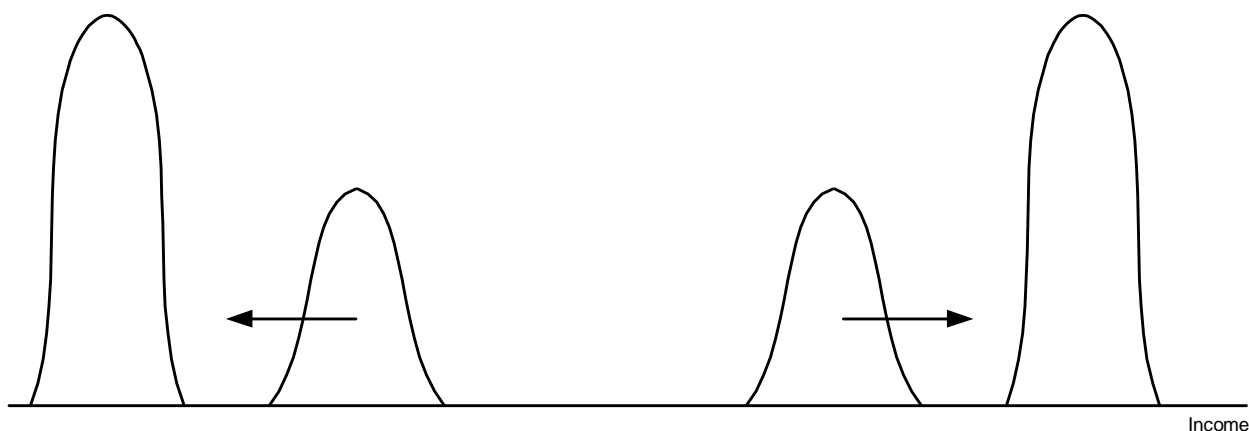


Figure 3: A “Symmetric Outward Slide” Must Raise Polarization.

which satisfy Axioms 1–4. The IA framework is, therefore, an essential part of the argument.

Notice that the characterized measures bear a superficial resemblance to the Gini coefficient. Indeed, if $\alpha = 0$, the measure *is* the Gini coefficient. However, our theorem ensures 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 densities. 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.⁹

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

⁹Indeed, 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 .

the decline in alienation to dominate the proceedings. This is tantamount to an upper bound on α .¹⁰

These restrictions — as well as a casual examination of the proof of Theorem 1 — make clear that the approach to our characterization, while superficially similar to Esteban and Ray [1994], is actually quite different. There we assumed discrete groups, which necessitated a very different set of axioms. It is comforting that both yield the same functional characterization, albeit with different numerical restrictions on the value of α .

We end this section with a discussion of our choice of basing identification on the point density. In section 3.1. we mentioned that we may more generally consider that individuals have 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 k be the parameter of the broadness of the sense of identification. Suppose that this just means that each individual x will consider an individual with income y to be at the point $(1 - k)x + ky$. It can be easily shown that the polarization measure resulting from this extended notion of identification is then proportional to our measure by the factor $k^{1-\alpha}$. Therefore, broadening the sense of identification simply amounts to a rescaling of the measure defined for the limit case when one is identified to individuals with exactly the same income.

It is also possible to directly base identification on the average density over a nondegenerate 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.

4. Estimation and Inference

We now turn to estimation issues regarding $P_\alpha(F)$, and associated questions of statistical inference.

4.1. Estimating $P_\alpha(F)$. The following rewriting of $P_\alpha(F)$ will be useful:

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

$$(12) \quad P_\alpha(F) = \int_y f(y)^\alpha g(y) dF(y) \equiv \int_y p_\alpha(y) dF(y),$$

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

¹⁰One might ask: why do the arguments in this paragraph and the one just before it lead to exactly the same 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.

Examining Observation 1, we may interpret $p_\alpha(y)$ as the contribution of individuals with income y to the overall index of polarization. As noted, $p_\alpha(y)$ has two multiplicative components. The first, $f(y)^\alpha$, captures the identification effect as we have already noted: the larger the value of α , the more urgent are “feelings of identification” in capturing polarization. The component $g(y)$ is the contribution of the alienation or distance effect in measuring polarization. When the identification effect is not important, that is, when $\alpha = 0$, we have that $p_0(y) = g(y)$ and $P_0 = \int g(y) dF(y)$ then equals 2μ times the Gini index.

Suppose that we wish to estimate $P_\alpha(F)$ using a random sample of n iid observations of income y_i , $i = 1, \dots, n$, drawn from the distribution $F(y)$ and ordered such that $y_1 \leq y_2 \leq \dots \leq y_n$. 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 $\hat{F}(y)$, by replacing $f(y)^\alpha$ by a suitable estimator $\hat{f}(y)^\alpha$ (to be examined below), and by replacing $g(y)$ by $\hat{g}(y)$. Hence, we have

$$(13) \quad P_\alpha(\hat{F}) = \int \hat{f}(y)^\alpha \hat{g}(y) d\hat{F}(y) = n^{-1} \sum_{i=1}^n \hat{f}(y_i)^\alpha \hat{g}(y_i),$$

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

$$(14) \quad \hat{F}(y_i) = \frac{1}{2} \left(\frac{(i-1)}{n} + \frac{(i)}{n} \right) = 0.5n^{-1} (2i-1)$$

and

$$(15) \quad \hat{\mu}^*(y_i) = n^{-1} \left(\sum_{j=1}^{i-1} y_j + \frac{i-(i-1)}{2} y_i \right),$$

and thus define $\hat{g}(y_i)$ as

$$(16) \quad \hat{g}(y_i) = \hat{\mu} + y_i (n^{-1} (2i-1) - 1) - n^{-1} \left(2 \sum_{j=1}^{i-1} y_j + y_i \right).$$

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

We have not yet discussed the estimator $\hat{f}(y)^\alpha$, but will do so presently. Observe, however, that an exact replication of the sample to the original sample should not change the value of the estimator $P_\alpha(\hat{F})$. Indeed, *presuming* that the estimators $\hat{f}(\cdot)^\alpha$ are invariant to sample size, this is indeed the case when formulae (13) and (16) are used. We record this formally as

Observation 2. Let $\mathbf{y} = (y_1, y_2, \dots, y_n)$ and $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{2n})$ be two vectors of sizes n and $2n$ respectively, ordered along increasing values of income. Suppose that for each $i \in \{1, \dots, n\}$, $y_i = \tilde{y}_{2i-1} = \tilde{y}_{2i}$ for all $i = 1, \dots, n$. Let $P_\alpha(F_{\mathbf{y}})$ be the polarization index defined by (13) and (16) for a vector of income \mathbf{y} . Then, provided that $f_{\mathbf{y}}(y_i) = f_{\tilde{\mathbf{y}}}(y_i)$ for $i = 1, \dots, n$, it must be that $P_\alpha(F_{\mathbf{y}}) = P_\alpha(F_{\tilde{\mathbf{y}}})$.

Remark. We may call this feature (sample) population-invariance.¹¹ When observations are weighted (or “grouped”), with w_i being the sampling weight on observation i and with $\bar{w} = \sum_{j=1}^n w_j$ being the sum of weights, a population-invariant

¹¹It is not to be confused with the *conceptual* discussion of what happens to polarization if the *true* population size is changed (and not that of the sample).

definition of $\widehat{g}(y_i)$ is then:

$$(17) \quad \widehat{g}(y_i) = \widehat{\mu} + y_i \left(\overline{w}^{-1} \left(2 \sum_{j=1}^i w_j - w_i \right) - 1 \right) - \overline{w}^{-1} \left(2 \sum_{j=1}^{i-1} w_j y_j + w_i y_i \right).$$

(16) is a special case of (17) obtained when $w_i = 1$ for all i . Setting $\alpha = 0$ in (13) and using (17) gives 2μ times the population-invariant Gini inequality index using a weighted sample. For analytical simplicity, however, we focus in what follows on the case of samples with unweighted iid observations.

4.2. $f(y_i)^\alpha$ and the Sampling Distribution of $P_\alpha(\widehat{F})$. For the measurement of polarization, it will be generally desirable to adjust our estimator of $f(y_i)^\alpha$ to sample size. The reason is that it will be statistically preferable to design our estimation of the identification effect to take into account the “quantity” of identifying information that exists in a sample, so as to minimize the sampling error of estimating the polarization indices.

To facilitate a more detailed discussion of this issue, we first decompose the estimator $P_\alpha(\widehat{F})$ across its separate sources of sampling variability:

$$(18) \quad \begin{aligned} P_\alpha(\widehat{F}) - P_\alpha(F) &= \int (\widehat{p}_\alpha(y) - p_\alpha(y)) dF(y) + \int p_\alpha(y) d(\widehat{F} - F)(y) \\ &+ \int (\widehat{p}_\alpha(y) - p_\alpha(y)) d(\widehat{F} - F)(y). \end{aligned}$$

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:

$$(19) \quad \begin{aligned} \widehat{p}_\alpha(y) - p_\alpha(y) &= \left(\widehat{f}(y)^\alpha - f(y)^\alpha \right) g(y) + f(y)^\alpha (\widehat{g}(y) - g(y)) \\ &+ \left(\widehat{f}(y)^\alpha - f(y)^\alpha \right) (\widehat{g}(y) - g(y)) \end{aligned}$$

As can be seen by inspection, $\widehat{g}(y) - g(y)$ 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 (19) 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 (18) is of order $o(n^{-1/2})$ and can also be ignored. Using all this information and combining (18) and (19), we see that for large n ,

$$(20) \quad P_\alpha(\widehat{F}) - P_\alpha(F) \cong \int \left(\widehat{f}(y)^\alpha - f(y)^\alpha \right) g(y) dF(y)$$

$$(21) \quad + \int f(y)^\alpha (\widehat{g}(y) - g(y)) dF(y)$$

$$(22) \quad + \int p_\alpha(y) d(\widehat{F} - F)(y).$$

The terms (21) and (22) are further developed in the proof of Theorem 2 in the appendix.

We thus turn to the estimation of $f(y)^\alpha$ in (20), which we propose to do nonparametrically using kernel density estimation. To this end, we make use of a kernel function $K(u)$, defined such that $\int_{-\infty}^{\infty} K(u) du = 1$ (this guarantees the desired

property that $\int_{-\infty}^{\infty} \hat{f}(y) dy = 1$) and $K(u) \geq 0$ (this guarantees that $\hat{f}(y) \geq 0$). It is also convenient to choose a kernel function that is symmetric around 0, with $\int uK(u)du = 0$ and $\int u^2K(u)du = \sigma_K^2 < \infty$. The estimator $\hat{f}(y)$ is then defined as

$$(23) \quad \hat{f}(y) \equiv n^{-1} \sum_{i=1}^n K_h(y - y_i),$$

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) for the kernel estimation procedure. For simplicity, we assume it to be invariant across y . We will see later how it should optimally be set as a function of the sample size, conditional on a choice of functional form for $K(u)$.

One kernel function that has nice continuity and differentiability properties is the Gaussian kernel, defined by

$$(24) \quad K(u) = (2\pi)^{-0.5} \exp^{-0.5u^2},$$

a form that we will use later for illustrative purposes.¹²

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

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 $n^{0.5} \left(P_\alpha(\hat{F}) - P_\alpha(F) \right)$ has a limiting normal distribution $N(0, V_\alpha)$, with*

$$(25) \quad V_\alpha = \mathop{\text{var}}_{f(y)} (a_\alpha(y)),$$

where

$$(26) \quad a_\alpha(y) = (1 + \alpha)p_\alpha(y) + y \int f(x)^\alpha dF(x) + 2 \int_y^\infty (x - y) f(x)^\alpha dF(x).$$

Observe that the assertion of Theorem 2 is distribution-free since everything in (25) can be estimated consistently *without* having to specify the population distribution from which the sample is drawn. $P_\alpha(\hat{F})$ 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 (25).

4.3. The Minimization of Sampling Error. In finite samples, however, $P_\alpha(\hat{F})$ 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 $P_\alpha(\hat{F})$ 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 can exploit this tradeoff to choose an "optimal" bandwidth for the estimation of $P_\alpha(\hat{F})$, 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

¹²Note that the Gaussian kernel has the property that $\sigma_K^2 = 1$.

(for a given h) the MSE into the sum of the squared bias and of the variance involved in estimating $P_\alpha(F)$:

$$(27) \quad \text{MSE}_h(P_\alpha(\hat{F})) = \left(\text{bias}_h \left(P_\alpha(\hat{F}) \right) \right)^2 + \text{var}_h \left(P_\alpha(\hat{F}) \right),$$

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

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

$$(28) \quad h^*(n) = \sqrt{-\frac{\text{cov}(a_\alpha(y), p''_\alpha(y))}{\alpha \sigma_K^2 \left(\int f''(y) p_\alpha(y) dy \right)^2} n^{-0.5} + O(n^{-1})}.$$

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 $a_\alpha(y)$ 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. On the one hand, 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 (28) to compute $h^*(n)$ (in the manner of Silverman (1986, p.45), for instance, for point density estimation). On other hand, we can estimate the terms in (28) directly from the empirical distribution, using an initial value of h to compute the $f(y)$ in the $a_\alpha(y)$ and $p_\alpha(y)$ functions. For both of these approaches (and particularly for the last one), expression (28) 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 (24)) is used to estimate $\hat{f}(y)$, h^* is approximately given by:

$$(29) \quad h^* \cong \frac{5.32 \sigma}{n^{0.5} \alpha^{0.65}}.$$

This formula works well with the normal distribution¹³ since it is rarely farther than 10% from the h^* that truly minimizes the MSE. It also seems to perform relatively well with other distributions, including the popular log-normal one. (29) is clearly also easily computed.

The use of this simple approximate rule may also be justified by the fact that the MSE of the polarization indices does not appear to be overly sensitive to the choice of the bandwidth h . This is shown for P_1 on Figure 6, again for the case of a normal distribution with $\sigma = 1$, and for $n = 1000$. Figure 6 shows the square root of the MSE for different choices of h , and is therefore an indication of the absolute amount by which we can expect estimates of polarization indices to differ from

¹³Extensive numerical simulations were made using various values of $n \geq 500$, σ and $\alpha = 0.25$ to 1. The results are available from the authors upon request.

the true population value¹⁴. Note that \sqrt{MSE} increases rather slowly as h moves away from its optimal value. Even with variations of ± 0.2 around h^* , for instance, \sqrt{MSE} barely changes.

5. An Illustration

We illustrate the above axiomatic and statistical results using data drawn from the Luxembourg Income Study (LIS) data sets¹⁵ on 21 countries for each of Wave 3 (1989–1992) and Wave 4 (1994–1997). The list of the countries, survey years and abbreviations is to be found in Table 1. [All figures and tables for this section are located at the end of the paper.] We use household disposable income (adult-equivalence scale defined as $h^{0.5}$, where h 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 suggested in the discussion of Theorem 1, the usual homogeneity-of-degree-zero property is imposed throughout by multiplying the indices $P_\alpha(F)$ by $\mu^{\alpha-1}$.

Tables 2 and 3 show estimates of the Gini ($P_{\alpha=0}$) and of two polarization indices (P_α for $\alpha = 0.25, 1$) in each of the 21 countries of each of the two waves, along with their asymptotic standard deviations in *italics*. The polarization indices are typically rather precisely estimated, with often only the third decimal of the estimators being subject to sampling variability. The rankings of the countries are very close for P_0 and $P_{0.25}$. But they differ considerably between P_0 and P_1 , and between $P_{0.25}$ 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. Hence, with these data at least, it seems that whether inequality comparisons resemble polarization comparisons depend on the differential ethical weight which is put on the *alienation versus* the *identification* effects.

To interpret the cross-country variability in the estimated value of these indices, it is useful to show graphically how each income level "contributes" to the total value of the indices. Recall equation (12): the polarization index $P_\alpha(F)$ is the integral of $p_\alpha(y)f(y)$ (a "polarization curve"), and the Gini index $P_{\alpha=0}(F)$ is the integral of $g(y)f(y)$ (an "alienation curve"). The difference between these two factors is the identification factor $f(y)^\alpha$. For illustrative purposes, these three factors (for $\alpha = 1$) are graphed on Figures 7 and 8 for the United States (1991) – us91 – and for the Czech Republic (1992) – cz92 – against income values normalized by the mean. The area underneath the $p_{\alpha=1}(y)f(y)$ and $g(y)f(y)$ curves then gives respectively the polarization index $P_{\alpha=1}$ and (twice) the Gini index. Figure 9 shows the Czech-US difference for each of these curves.

By definition, the area underneath each of the $f(y)$ density curves gives 1, and the area covered by their difference gives 0. From Figure 9, it is clear why Gini inequality is lower in cz92 than in us91: the us91 alienation curve (the curve with the rectangular dots) is almost everywhere larger than the cz92 alienation curve.

¹⁴Note that around the optimal value of h (which is about 0.16), the absolute value of the bias is of the order of one tenth of \sqrt{MSE} – in the case of this example at least, its computation could therefore be safely avoided.

¹⁵<http://lissy.ceps.lu> for detailed information on the structure of these data.

The area defined by the polarization curve (the curve with the circular dots) is, however, slightly positive, and this explains why polarization is slightly larger in cz92.

The reasons for these differences can be seen from Figures 7 and 8. Roughly speaking, for a given level of inequality, a country will exhibit a large level of polarization if there is a strong correlation between the *inequality* or *alienation* component $g(y)$ and the *identification* one $f(y)^\alpha$. *Ceteris paribus*, such a strong correlation will lead to a large area underneath the $p_\alpha(y)f(y)$ curve. An example of such a sharp correlation can be seen for the Czech Republic in 1992 (Figure 8). The density is concentrated where alienation is felt most: in this example, this is around a normalized income of about 0.75. A converse example of a weaker correlation is that of the USA in both Wave 3 and Wave 4. In Figure 7, the density is relatively flat, and not obviously concentrated where alienation is most felt, that is, at normalized incomes of about 0.4.

Differences in alienation-identification correlations thus explain why the inequality and polarization rankings of countries differ sometimes very significantly in Tables 2 and 3. Some countries such as Finland, Sweden and Denmark rank low both in terms in inequality and polarization (and for both waves). Other countries show low inequality but relatively high polarization, while others exhibit the reverse relative rankings. Some countries, most strikingly Russia and Mexico, finally rank consistently high both in terms in inequality and polarization.

Tables 4 to 6 show which ones of these cross-country rankings are statistically significant and can therefore be reasonably attributed to true population differences in inequality and polarization. The results are for Wave 3 countries, and for $\alpha = 0, 0.25$ and 1. The Tables show p -values of tests that the countries listed on the first row show more inequality (Table 4) or polarization (Tables 5 and 6) than countries on the first column. Roughly speaking, these p -values indicate the probability that an error is made when one rejects the null hypotheses that countries on the first row do not have a larger P_α than countries on the first column, in favor of the alternative hypotheses that P_α is indeed greater for the countries on the first row. More formally, such p -values are the maximal test sizes that will lead to the rejection of the above null hypotheses. Using a conventional test size of 5%, it can be seen that many (all those with a *, viz, around 90%) of the possible cross-country comparisons are statistically significant. This is true for all three values of α .

As mentioned above, there are many country pairs whose polarization ordering sometimes differs from their Gini ordering, and whose polarization ordering also varies with the choice of α . One way to shed light on this issue is to estimate the range of α values for which one country has a higher P_α index than another. To illustrate this, take the case of us91, cz92 and uk91. Ignoring sampling variability, cz92 has a lower P_α index than us91 for all $\alpha < 0.98$, including the Gini index. In such a case, it would therefore seem that inequality and polarization rankings *almost* agree, although not quite completely. uk91 has, however, a higher polarization index than us91 for all (and only for) $\alpha > 0.33$, and thus has a lower Gini index than us91. Clearly then, the Gini ranking of uk91 and us91 will differ from their ranking according to most of this paper's polarization indices. In these situations, incorporating identification effects would therefore appear to change the distributive picture substantially.

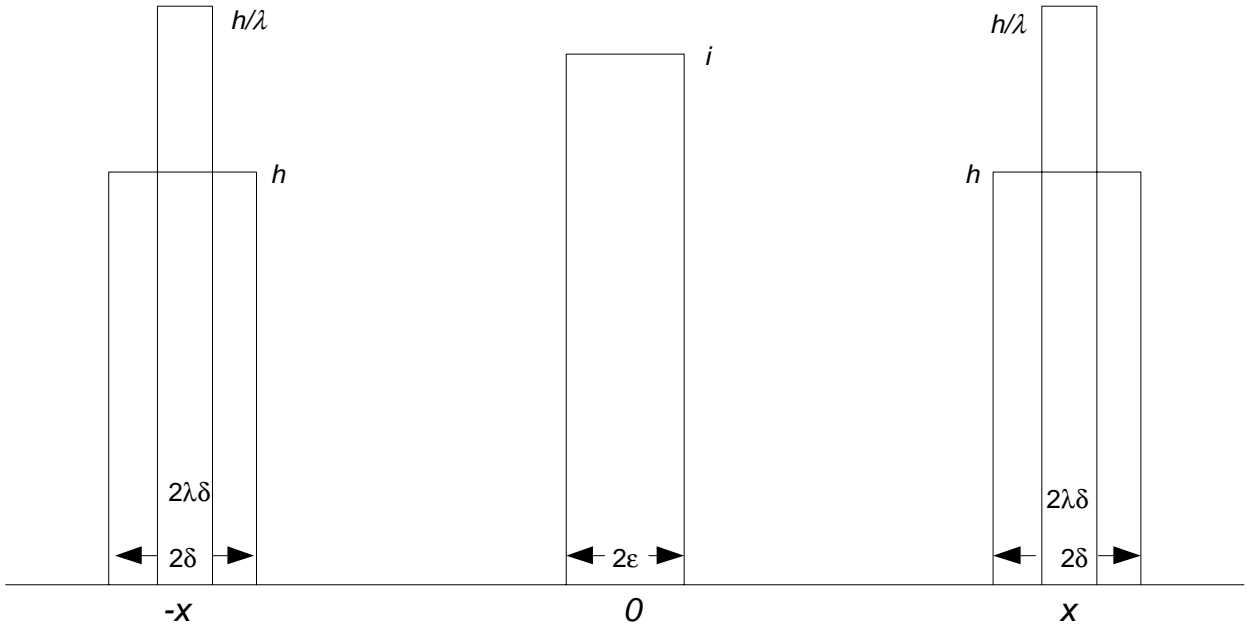


Figure 4:

6. Proofs

Proof of Theorem 1. In the first half of the proof, we show that axioms 1–4 imply (11).

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

$$(30) \quad g(x) \geq \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} g(y) dy.$$

Then g must be a concave function.

Proof. This is a well-known implication of Jensen's characterization of concave functions. ■

In what follows, keep in mind that the basic structure of our measure only considers income *differences* across people, and not the incomes per se. Therefore we may slide any distribution to the left or right as we please, without disturbing the analysis (even negative incomes may be considered when these are expositively convenient).

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

Proof. Fix $x > 0$, some $i > 0$, and some value of $\delta \in (0, x)$. Consider the following specialization of the setting of Axiom 2. We take three basic densities as in that Axiom (see also Figure 1) but specialize as shown in Figure 4; each is a transform

of a uniform kernel. 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 . In the sequel, we shall vary ϵ and h but to make sure that Axiom 2 applies, we choose $\epsilon > 0$ such that $\delta + \epsilon < x$. It is easy to check that a λ -squeeze of the side densities simply implies that the base of the rectangle is contracted to a width $2\lambda\delta$ (keeping the centering unchanged), while the height is raised to h/λ . See Figure 4. For each λ , we may decompose the polarization measure (9) into five distinct components. First, there is the “internal polarization” of the middle rectangle, call it P_m . This component is unchanged as we change λ so there will be no need to explicitly calculate it. Next, there is the “internal polarization” of each of the side rectangles, call it P_s . Third, there is the total effective antagonism felt by inhabitants of the middle towards each side density. Call this A_{ms} . Fourth, there is the total effective antagonism felt by inhabitants of each side towards the middle. Call this A_{sm} . Finally, there is the total effective antagonism felt by inhabitants of one side towards the other side. Call this A_{ss} . Observe that each of these last four terms appear twice, so that (writing everything as a function of λ),

$$(31) \quad P(\lambda) = P_m + 2P_s(\lambda) + 2A_{ms}(\lambda) + 2A_{sm}(\lambda) + 2A_{ss}(\lambda),$$

Now we compute the terms on the right hand side of (31). First,

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

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,

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

Third,

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

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

$$(32) \quad \liminf_{h \rightarrow 0} \frac{P(\lambda) - P(1)}{2h} \geq 0.$$

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, (32) and the calculations above must jointly imply that

$$(33) \quad \frac{1}{\lambda} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda\delta}^{x+\lambda\delta} T(i, b' - b) db' db \geq \int_{-\epsilon}^{\epsilon} \int_{x-\delta}^{x+\delta} T(i, b' - b) db' db,$$

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 (33) holding as $\lambda \rightarrow 0$. Performing the necessary

calculations, we may conclude that

$$(34) \quad \frac{1}{\epsilon} \int_{-\epsilon}^{\epsilon} T(i, x-b) db \geq \frac{1}{\epsilon} \int_{-\epsilon}^{\epsilon} \int_{x-\delta}^{x+\delta} T(i, b'-b) db' db$$

for every $\epsilon \in (0, x-\delta)$. Finally, take ϵ to zero in (34). This allows us to deduce that

$$(35) \quad T(i, x) \geq \int_{x-\delta}^{x+\delta} T(i, b') db'.$$

As (35) 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 $i > 0$. ■

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

$$(36) \quad g(a + \Delta) - g(a) \geq g(a') - g(a' - \Delta)$$

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

Proof. Given the concavity of g , it is easy to see that

$$g(a + \Delta) - g(a) \leq g(a') - g(a' - \Delta)$$

for all $a > a' \geq \Delta > 0$. Combining this information with (36), we may conclude that for each a and a' with $a > a' > 0$, there exists $\bar{\Delta} > 0$ such that

$$g(a + \Delta) - g(a) = g(a') - g(a' - \Delta)$$

for all $\Delta \in (0, \bar{\Delta})$. This, coupled with the premises that g is concave and $g(0) = 0$, shows that g is linear. ■

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

Proof. Fix numbers a and a' with $a > a' > 0$, and $i > 0$. Consider the following specialization of Axiom 3: take four basic densities as in that Axiom (see also Figure 3) but specialize as shown in Figure 5; each is a transform of a uniform kernel. 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 . In the sequel, we shall vary δ, ϵ and h but to make sure that the basic densities have disjoint support, we restrict ourselves to values of δ and ϵ such that $\epsilon < x$ and $\delta + \epsilon < y - x - \bar{\Delta}$ for some $\bar{\Delta} > 0$. For convenience, the rectangles have been numbered 1, 2, 3 and 4 for use below. The exercise that we perform is to increase x by the small amount Δ , where $0 < \Delta < \bar{\Delta}$, as defined above. Given this configuration, we may decompose the polarization measure (9) into several distinct components. First, there is 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. Next, there is the total effective antagonism felt by inhabitants of each rectangle towards another; call this $A_{jk}(x)$, where j is the "origin" rectangle and k is the "destination" rectangle. [We emphasize the dependence on x , which is the parameter to be varied.] Thus total polarization $P(x)$, again written explicitly as a function of x , is given by

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

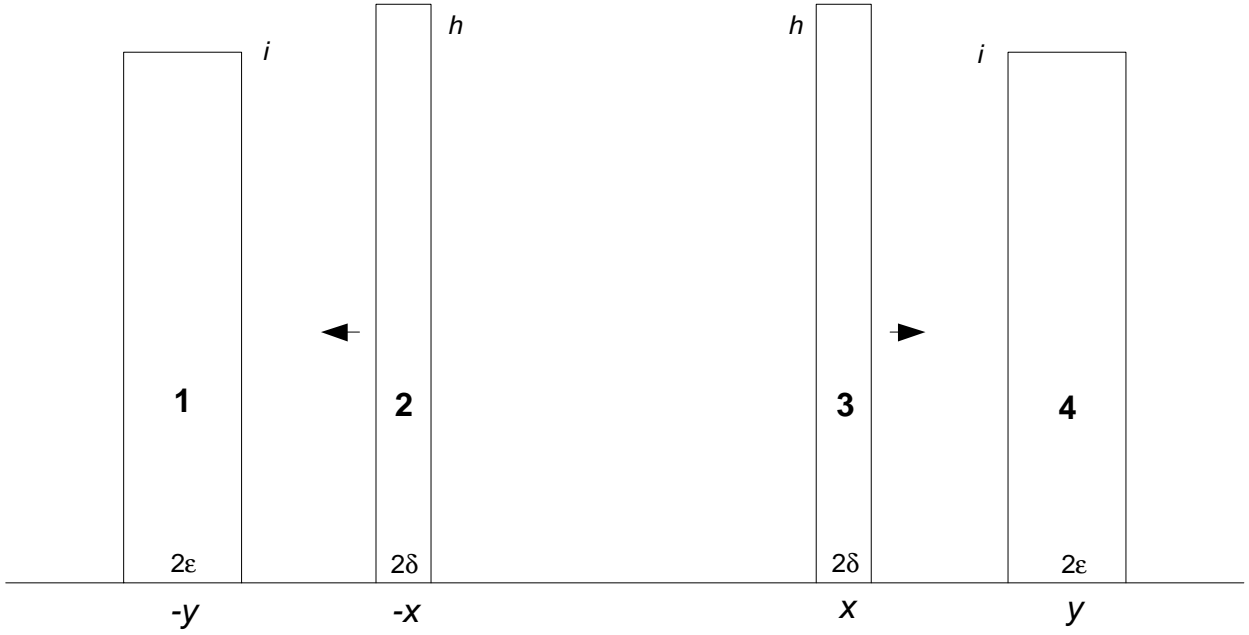


Figure 5:

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

$$A_{12}(x) = \int_{-y-\epsilon}^{-y+\epsilon} \int_{-x-\delta}^{-x+\delta} T(i, b' - b) i h d b' d b,$$

$$A_{13}(x) = \int_{-y-\epsilon}^{-y+\epsilon} \int_{x-\delta}^{x+\delta} T(i, b' - b) i h d b' d b,$$

$$A_{21}(x) = \int_{-x-\delta}^{-x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') i h d b' d b,$$

$$A_{31}(x) = \int_{x-\delta}^{x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') i h d b' d b,$$

and

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

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

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

Using this information along with the computations for $P(x)$ and the various $A_{jk}(x)$'s, we see that

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

where in arriving at this inequality, we have carried out some elementary substitution of variables and transposition of terms. 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 $T(i, \cdot)$ must be linear for every $i > 0$ since $T(0, a) := 0$. But this only means that 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 . ■

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

Proof. As a preliminary step, observe that

$$(37) \quad \phi(i) > 0 \text{ whenever } i > 0.$$

[For, if this were false for some i , Axiom 3 would fail for configurations constructed from rectangular basic densities of equal height i .] Our first objective is to prove that ϕ must satisfy the fundamental Cauchy equation

$$(38) \quad \phi(p)\phi(p') = \phi(pp')\phi(1)$$

for every strictly positive p and p' . To this end, fix p and p' and define $r \equiv pp'$. In what follows, we assume that $p \geq r$. [If $r \geq p$, simply permute p and r in the argument below.] Consider the following configuration. There are 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 is any strictly positive number, soon to be made arbitrarily small). It is easy to see that the polarization of this configuration, P , is given by

$$\begin{aligned} P &= ph[\phi(p) + \phi(h)] \left\{ \int_{-\epsilon}^{\epsilon} \int_{1-\epsilon}^{1+\epsilon} (b' - b) db' db \right\} \\ &\quad + [p^2\phi(p) + h^2\phi(h)] \left\{ \int_{-\epsilon}^{\epsilon} \int_{-\epsilon}^{\epsilon} |b' - b| db' db \right\} \\ (39) \quad &= 4\epsilon^2 ph[\phi(p) + \phi(h)] + \frac{8\epsilon^3}{3} [p^2\phi(p) + h^2\phi(h)], \end{aligned}$$

where the first equality invokes Lemma 4 and both equalities use routine computations. Now change the height of the first rectangle to r . Using (37) and the provisional assumption that $p \geq r$, it is easy to see that for each ϵ , there must exist a (unique) height for the second rectangle — call it $h(\epsilon)$, such that the polarizations

of the two configurations are equated. Invoking (39), we equivalently choose $h(\epsilon)$ such that

$$(40) \quad \begin{aligned} & ph[\phi(p) + \phi(h)] + \frac{2\epsilon}{3}[p^2\phi(p) + h^2\phi(h)] \\ &= rh(\epsilon)[\phi(r) + \phi(h(\epsilon))] + \frac{2\epsilon}{3}[r^2\phi(r) + h(\epsilon)^2\phi(h(\epsilon))]. \end{aligned}$$

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

$$(41) \quad \begin{aligned} & \lambda^2 ph[\phi(\lambda p) + \phi(\lambda h)] + \frac{2\epsilon}{3}[(\lambda p)^2\phi(\lambda p) + (\lambda h)^2\phi(\lambda h)] \\ &= \lambda^2 rh(\epsilon)[\phi(\lambda r) + \phi(\lambda h(\epsilon))] + \frac{2\epsilon}{3}[(\lambda r)^2\phi(\lambda r) + [\lambda h(\epsilon)]^2\phi(\lambda h(\epsilon))]. \end{aligned}$$

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 (40) and (41) to conclude that

$$(42) \quad ph[\phi(p) + \phi(h)] = rh'[\phi(r) + \phi(h')]$$

and

$$(43) \quad \lambda^2 ph[\phi(\lambda p) + \phi(\lambda h)] = \lambda^2 rh'[\phi(\lambda r) + \phi(\lambda h')].$$

Combining (42) and (43), we see that

$$(44) \quad \frac{\phi(p) + \phi(h)}{\phi(\lambda p) + \phi(\lambda h)} = \frac{\phi(r) + \phi(h')}{\phi(\lambda r) + \phi(\lambda h')}.$$

Taking limits in (44) as $h \rightarrow 0$ and noting that $h' \rightarrow 0$ as a result (examine (42) to confirm this), we have for all $\lambda > 0$,

$$(45) \quad \frac{\phi(p)}{\phi(\lambda p)} = \frac{\phi(r)}{\phi(\lambda r)}.$$

Put $\lambda = 1/p$ and recall that $r = pp'$. Then (45) yields the required Cauchy equation (38). To complete the proof, recall that ϕ is continuous and that (37) holds. The class of solutions to (38) (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. First notice that each basic density f with mass p , support $[a, b]$ and mean μ may be connected to its kernel — call it f^* — by means of three numbers. First, we *slide* the density so that it begins at 0; this amounts to a slide of a to the left. The new mean is now $m \equiv \mu - a$. Second, we *income-scale* the density so as to change its mean from $m = \mu - a$ to 1. Finally, we *population-scale* to change the overall mass of the density from p to unity.

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 kernel of f . Then, if f^λ denotes some λ -squeeze of f ,*

$$(46) \quad P(F^\lambda) = 4kp^{2+\alpha}(m\lambda)^{1-\alpha} \int_0^1 f^*(x)^{1+\alpha} \left\{ \int_0^1 f^*(y)(1-y)dy + \int_x^1 f^*(y)(y-x)dy \right\} dx$$

for some constant $k > 0$.

Proof. Let f be given as in the statement of the lemma. Recall that a slide of the entire distribution 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 (11),

$$(47) \quad P(F) = k \int \int f(x)^{1+\alpha} f(y) |y - x| dy dx$$

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

$$(48) \quad \begin{aligned} P(F) &= 2k \int_0^m \int_0^{2m} f(x')^{1+\alpha} f(y') |x' - y'| dy' dx' \\ &= 2k \int_0^m f(x')^{1+\alpha} \left\{ \int_0^{x'} f(y') (x' - y') dy' + \int_{x'}^m f(y') (y' - x') dy' \right. \\ &\quad \left. + \int_m^{2m} f(y') (y' - x') dy' \right\} dx'. \end{aligned}$$

Examine the very last term in (48). 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 (48), and manipulating terms, we obtain

$$(49) \quad P(F) = 4k \int_0^m f(x')^{1+\alpha} \left\{ \int_0^m f(y') (m - y') dy' + \int_{x'}^m f(y') (y' - x') dy' \right\} dx'.$$

Now suppose that f^λ is a λ -squeeze of f . Note that (49) holds just as readily for f^λ as for f . Therefore, using the expression for f given in (10), we see that

$$(49) \quad \begin{aligned} P(F^\lambda) &= 4k\lambda^{-(2+\alpha)} \int_{(1-\lambda)m}^m f\left(\frac{x' - (1-\lambda)m}{\lambda}\right)^{1+\alpha} \left\{ \int_{(1-\lambda)m}^m f\left(\frac{y' - (1-\lambda)m}{\lambda}\right) (m - y') dy' \right. \\ &\quad \left. + \int_{x'}^m f\left(\frac{y' - (1-\lambda)m}{\lambda}\right) (y' - x') dy' \right\} dx'. \end{aligned}$$

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 kernel 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 (46). ■

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 kernel of f . Then for any λ -squeeze f^λ of f ,

$$(50) \quad A(f^\lambda, g) = 2dkp^{1+\alpha}q(m\lambda)^{-\alpha} \int_0^1 f^*(x)^{1+\alpha} dx,$$

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

Proof. To begin with, ignore the λ -squeeze. Notice that there is no loss of generality in assuming that every income under g dominates every income under f . It also makes no difference to polarization whether or not we slide the entire configuration to the left or right. Therefore we may suppose that f has support $[0, 2m]$ (with mean m) and g has support $[d, d + 2m]$ (where obviously we must have $d \geq 2m$ for the disjoint support assumption to make sense). Because (47) is true, it must be that

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

where the third equality exploits the symmetry of g ,¹⁶ the fourth equality uses the fact that $\int_d^{d+m} g(y) = q/2$, and the final equality uses the symmetry of f .¹⁷ 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 kernel 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

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

¹⁶That is, for each $y \in [d, d + m]$, $g(y) = g(d + 2m - (y - d)) = g(2d + 2m - y)$. Moreover, $[y - x] + [(2d + 2m - y) - x] = 2(d + m - x)$.

¹⁷That is, for each $x \in [0, m]$, $f(x) = f(2m - x)$. Moreover, $[m + d - x] + [m + d - (2m - x)] = 2d$.

Finally, make the change of variables $x = x''/m$. Noting that $f^*(z) = mh(mz)$, we get (50). ■

Lemma 8. Define, for any kernel f and $\alpha > 0$,

$$(51) \quad \psi(f, \alpha) \equiv \frac{\int_0^1 f(x)^{1+\alpha} dx}{\int_0^1 f(x)^{1+\alpha} \left\{ \int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy \right\} dx}.$$

Then — for any $\alpha > 0$ — $\psi(f, \alpha)$ attains its minimum value when f is the uniform kernel, 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

$$(52) \quad L(x) \equiv \int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy$$

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 kernel, the weights must be nondecreasing in x . It follows that

$$(53) \quad \zeta(f, \alpha) \leq \int_0^1 L(x) dx.$$

Now

$$(54) \quad \begin{aligned} L(x) &= \int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy \\ &= \int_0^1 f(y)(1-x)dy + \int_0^x f(y)(x-y)dy \\ &= \frac{1-x}{2} + \int_0^x f(y)(x-y)dy. \end{aligned}$$

Because $f(x)$ is nondecreasing and integrates to 1/2 on $[0, 1]$, it must be the case that $\int_0^x f(y)(x-y)dy \leq \int_0^x (x-y)/2 dy$ for all $x \leq 1$. Using this information in (54) and combining it with (53),

$$(55) \quad \begin{aligned} \zeta(f, \alpha) &\leq \int_0^1 \left[\frac{1-x}{2} + \int_0^x \frac{x-y}{2} dy \right] dx \\ &= \int_0^1 \left[\int_0^1 \left[\frac{1-y}{2} \right] dy + \int_x^1 \left[\frac{y-x}{2} \right] dy \right] dx \\ &= \zeta(u, \alpha), \end{aligned}$$

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

Lemma 9. Given that $P(F)$ is of the form (47), Axiom 1 is satisfied if and only if $\alpha \leq 1$.

Proof. Simply inspect (46). ■

Lemma 10. *Given that $P(F)$ is of the form (47), 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 kernel; 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 kernel 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 (47) 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

$$(56) \quad P(\lambda) = P_m + 2P_s(\lambda) + 2A_{ms}(\lambda) + 2A_{sm}(\lambda) + 2A_{ss}(\lambda).$$

Compute these terms. For brevity, define for any kernel 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 (56), we may conclude that

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

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 (51) 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,

$$(57) \quad 2\lambda^{1-\alpha} + \frac{d}{m}\psi(f^*, \alpha)\lambda^{-\alpha} \left[\frac{q}{p} + 2 \right]$$

must be nonincreasing in λ over $(0, 1]$. An examination of the expression in (57) 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 kernel f^* ,

$$(58) \quad \lambda^{1-\alpha} + \psi(f^*, \alpha)\lambda^{-\alpha}$$

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

$$(59) \quad \alpha \geq \frac{1}{1 + \psi(f^*, \alpha)}.$$

Therefore, to find the necessary and sufficient bound on α (uniform over all kernels), we need to minimize $\psi(f^*, \alpha)$ by choice of f^* , subject to the condition that f^* be a kernel. By Lemma 8, this minimum value is 3. Using this information in (59), we are done. ■

Lemma 11. *Given that $P(F)$ is of the form (47), 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 5. 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,

$$(60) \quad P(x) - P(0) = 2\{[A_{12}(x) + A_{13}(x)] - [A_{12}(0) + A_{13}(0)]\} + [A_{23}(x) - A_{23}(0)]$$

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

Proof of Observation 1. First note that $|x - y| = x + y - 2 \min(x, y)$. Hence, by (11),

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

To prove (12), note that

$$(61) \quad \int_x \int_y x f(y)^\alpha dF(y) dF(x) = \mu \int_y f(y)^\alpha dF(y)$$

and that

$$(62) \quad \int_x \int_y f(y)^\alpha \min(x, y) dF(y) dF(x) = \int_x \int_{y=-\infty}^{y=x} y f(y)^\alpha dF(y) dF(x) + \int_x \int_{y=x}^{\infty} x f(y)^\alpha dF(y) dF(x).$$

The first term in (62) can be integrated by parts over x :

$$(63) \quad \begin{aligned} & \int_{y=-\infty}^{y=x} y f(y)^\alpha dF(y) F(x) \Big|_{-\infty}^{\infty} - \int x f(x)^\alpha F(x) dF(x) \\ &= \int y f(y)^\alpha dF(y) - \int x f(x)^\alpha F(x) dF(x) \\ &= \int y f(y)^\alpha (1 - F(y)) dF(y). \end{aligned}$$

The last term in (62) can also be integrated by parts over x as follows:

$$(64) \quad \begin{aligned} \int_x \int_{y=x}^{\infty} x f(y)^\alpha dF(y) dF(x) &= \int_x \int_{y=x}^{\infty} f(y)^\alpha dF(y) x dF(x) \\ &= \mu^*(x) \int_{y=x}^{\infty} f(y)^\alpha dF(y) \Big|_{x=-\infty}^{x=\infty} + \int_x \mu^*(x) f(x)^\alpha dF(x) \\ &= \int_y \mu^*(y) f(y)^\alpha dF(y), \end{aligned}$$

where $\mu^*(x) = \int_{-\infty}^x z dF(z)$ is a partial mean. Adding terms yields (12), and completes the proof. ■

Proof of Observation 2. It will be enough to show that $2g_y(y_i) = g_{\tilde{y}}(\tilde{y}_{2i-1}) + g_{\tilde{y}}(\tilde{y}_{2i})$ since we have assumed that $f_y(y_i) = f_{\tilde{y}}(\tilde{y}_{2i-1}) = f_{\tilde{y}}(\tilde{y}_{2i})$ for all $i = 1, \dots, n$.

Clearly, $\mu_{\mathbf{y}} = \mu_{\tilde{\mathbf{y}}}$. Note also that $g_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1})$ can be expressed as

$$(65) \quad g_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1}) = \mu_{\mathbf{y}} + y_i \left((2n)^{-1} (2(2i-1) - 1) - 1 \right) - (2n)^{-1} \left(2 \sum_{j=1}^{2i-2} \tilde{y}_j + \tilde{y}_{2i-1} \right).$$

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

$$(66) \quad g_{\tilde{\mathbf{y}}}(\tilde{y}_{2i}) = \mu_{\mathbf{y}} + y_i \left((2n)^{-1} (2(2i) - 1) - 1 \right) - (2n)^{-1} \left(2 \sum_{j=1}^{2i-1} \tilde{y}_j + \tilde{y}_{2i} \right).$$

Summing (65) and (66), we find

$$(67) \quad \begin{aligned} g_{\tilde{\mathbf{y}}}(\tilde{y}_{2i-1}) + g_{\tilde{\mathbf{y}}}(\tilde{y}_{2i}) &= 2 \left(\mu_{\mathbf{y}} + y_i (n^{-1}(2i-1) - 1) - n^{-1} \left(2 \sum_{j=1}^{i-1} y_j + y_i \right) \right) \\ &= 2g_{\mathbf{y}}(y_i). \end{aligned}$$

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

Proof of Theorem 2. Consider first (20). Note that

$$(68) \quad \begin{aligned} \int \left(\hat{f}(y)^{\alpha} - f(y)^{\alpha} \right) g(y) dF(y) &\cong \int \alpha f(y)^{\alpha-1} \left(\hat{f}(y) - f(y) \right) g(y) dF(y) \\ &= \alpha \int p_{\alpha-1}(y) n^{-1} \sum_{i=1}^n K_h(y - y_i) dF(y) - \alpha \int p_{\alpha}(y) dF(y) \\ &= \alpha n^{-1} \sum_{i=1}^n \int p_{\alpha-1}(y) K_h(y - y_i) dF(y) - \alpha \int p_{\alpha}(y) dF(y). \end{aligned}$$

Taking $h \rightarrow 0$ as $n \rightarrow \infty$, and recalling that $\int K_h(y - y_i) dy = 1$, the first term in (68) tends asymptotically to

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

Thus, we can rewrite the term on the right-hand side of (20) as

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

Now turn to (21). Let I be an indicator function that equals 1 if its argument is true and 0 otherwise. We find:

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

Now consider (22):

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

Collecting and summarizing terms, we obtain:

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

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

Proof of Theorem 3. Using (20)–(22), we may write $\text{bias}_n(\widehat{F}_\alpha) = \mathbf{E} [P_\alpha(\widehat{F}) - P_\alpha(F)]$ as:

$$\begin{aligned}
\mathbf{E} [P_\alpha(\widehat{F}) - P_\alpha(F)] &\cong \int \mathbf{E} [\widehat{f}(y)^\alpha - f(y)^\alpha] g(y) dF(y) \\
&\quad + \int f(y)^\alpha \mathbf{E} [\widehat{g}(y) - g(y)] dF(y) + \int p_\alpha(y) d\mathbf{E} [\widehat{F} - F](y) \\
(70) \qquad \qquad \qquad &= \int \mathbf{E} [\widehat{f}(y)^\alpha - f(y)^\alpha] g(y) dF(y),
\end{aligned}$$

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

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

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

$$(71) \quad 0.5h^2\sigma_K^2 f''(y),$$

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

$$(72) \quad \mathbf{E}[P_\alpha(\hat{F}) - P_\alpha(F)] \cong 0.5\alpha\sigma_K^2 h^2 \int f''(y)p_\alpha(y)dy = O(h^2).$$

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 $\text{var}_h(P_\alpha(\hat{F}))$, we first reconsider the first term in (68), which is the dominant term through which the choice of h influences $\text{var}(P_\alpha(\hat{F}))$. We may write this as follows:

$$(73) \quad \begin{aligned} \alpha n^{-1} \sum_{i=1}^n \int p_\alpha(y) K_h(y - y_i) dy &= \alpha n^{-1} \sum_{i=1}^n \int p_\alpha(y_i - ht) K(t) dt \\ &\cong \alpha n^{-1} \sum_{i=1}^n \int K(t) (p_\alpha(y_i) - htp'_\alpha(y_i) + 0.5h^2t^2p''_\alpha(y_i)) dt \\ &= \alpha n^{-1} \sum_{i=1}^n (p_\alpha(y_i) + 0.5\sigma_K^2 h^2 p''_\alpha(y_i)), \end{aligned}$$

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 (73) and (25) to incorporate a finite-sample correction for the role of h in the variance of \hat{F}_α , we can write:

$$\text{var}_h(P_\alpha(\hat{F})) = n^{-1} \text{var}_{f(y)}(0.5\alpha\sigma_K^2 h^2 p''_\alpha(y) + a_\alpha(y)) = O(n^{-1}).$$

For small h , the impact of h on the finite sample variance comes predominantly from the covariance between $a_\alpha(y)$ and $p''_\alpha(y)$ since $\text{var}(0.5\alpha\sigma_K^2 h^2 p''_\alpha(y))$ 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 $\text{var}_h(P_\alpha(\hat{F}))$ since this will tend to level the distribution of $0.5\alpha\sigma_K^2 h^2 p''_\alpha(y) + a_\alpha(y)$, which is the random variable whose variance determines the sampling variance of $P_\alpha(\hat{F})$. Combining squared-bias and variance into (27), we obtain:

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

$h^*(n)$ 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:

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

Since $h^*(n) > 0$ in finite samples, we may divide (74) by h , and then find $h^*(n)$ by setting this first-order condition to 0. This yields:

$$(75) \quad h^*(n)^2 = - \frac{n^{-1} \text{cov}(a_\alpha(y), p_\alpha''(y))}{\alpha \sigma_K^2 \left(\left(\int f''(y) p_\alpha(y) dy \right)^2 - 0.5 n^{-1} \text{var}(p_\alpha''(y) p_\alpha(y)) \right)}$$

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

$$(76) \quad h^*(n) = \sqrt{- \frac{\text{cov}(a_\alpha(y), p_\alpha''(y))}{\alpha \sigma_K^2 \left(\int f''(y) p_\alpha(y) dy \right)^2} n^{-0.5} + O(n^{-1})}$$

This completes the proof. ■

References

- Alesina, A. and E. Spolaore (1997) "On the Number and Size of Nations," *Quarterly Journal of Economics* **113**, 1027–1056.
- Aghevli, B.B. and F. Mehran (1981), "Optimal Grouping of Income Distribution Data," *Journal of the American Statistical Association* **76**, 22–26.
- 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.
- Davidson, R. and J.Y. Duclos (2000), "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, **68**, 1435–1465.
- Davies, J.B. and A.F. Shorrocks (1989), "Optimal Grouping of Income and Wealth Data," *Journal of Econometrics* **42**, 97–108.

- D'Ambrosio, C. and E. Wolff (2001), "Is Wealth Becoming More Polarized in the United States?" Working Paper 330, Levy Economics Institute, Bard College.
- Esteban, J. and D. Ray (1991) "On the Measurement of Polarization," Boston University, Institute for Economic Development, Working Paper 18.
- Esteban, J. and D. Ray (1994) "On the Measurement of Polarization," *Econometrica* **62**, 819–852.
- Esteban, J. and D. Ray (1999) "Conflict and Distribution," *Journal of Economic Theory* **87**, 379–415.
- 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.
- 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.
- Quah, D. (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.
- 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.
- Wolfson, M.C. (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 1: LIS country codes

- as = Australia
- be = Belgium
- cn = Canada
- cz = Czech Republic
- dk = Denmark
- fi = Finland
- fr = France
- ge = Germany
- hu = Hungary
- is = Israel
- it = Italy
- lx = Luxembourg
- mx = Mexico
- nl = Netherlands
- nw = Norway
- pl = Poland
- rc = Republic of China and Taiwan
- ru = Russia
- sw = Sweden
- uk = United Kingdom
- us = United States

Table 2: Ranking from LIS' Wave 3

LIS Country	Index <i>StdDev</i>	Ranking	Index <i>StdDev</i>	Ranking	Index <i>StdDev</i>	Ranking
$\alpha =$	0		0.25		1	
cz92	0.2082 <i>0.0023</i>	1	0.1770 <i>0.0014</i>	1	0.1574 <i>0.0012</i>	11
fi91	0.2086 <i>0.0017</i>	2	0.1785 <i>0.0011</i>	2	0.1437 <i>0.0005</i>	1
be92	0.2236 <i>0.0028</i>	3	0.1903 <i>0.0018</i>	4	0.1489 <i>0.0011</i>	3
sw92	0.2267 <i>0.0019</i>	4	0.1892 <i>0.0012</i>	3	0.1459 <i>0.0006</i>	2
nw91	0.2315 <i>0.0029</i>	5	0.1925 <i>0.0018</i>	5	0.1504 <i>0.0011</i>	5
dk92	0.2367 <i>0.0026</i>	6	0.1970 <i>0.0015</i>	6	0.1502 <i>0.0011</i>	4
lx91	0.2389 <i>0.0051</i>	7	0.2013 <i>0.0033</i>	7	0.1569 <i>0.0023</i>	10
ge89	0.2479 <i>0.0049</i>	8	0.2031 <i>0.0029</i>	8	0.1537 <i>0.0021</i>	7
nl91	0.2633 <i>0.0054</i>	9	0.2127 <i>0.0032</i>	9	0.1591 <i>0.0025</i>	16
rc91	0.2708 <i>0.0019</i>	10	0.2193 <i>0.0012</i>	10	0.1603 <i>0.0009</i>	17
pl92	0.2737 <i>0.0032</i>	11	0.2198 <i>0.0019</i>	11	0.1575 <i>0.0013</i>	13
fr89	0.2815 <i>0.0033</i>	12	0.2233 <i>0.0019</i>	12	0.1577 <i>0.0014</i>	14
hu91	0.2828 <i>0.0066</i>	13	0.2237 <i>0.0040</i>	13	0.1582 <i>0.0027</i>	15
it91	0.2887 <i>0.0028</i>	14	0.2315 <i>0.0017</i>	15	0.1575 <i>0.0012</i>	12
cn91	0.2891 <i>0.0018</i>	15	0.2312 <i>0.0012</i>	14	0.1521 <i>0.0006</i>	6
is92	0.3055 <i>0.0036</i>	16	0.2427 <i>0.0021</i>	17	0.1625 <i>0.0015</i>	18
as89	0.3084 <i>0.0020</i>	17	0.2427 <i>0.0012</i>	16	0.1547 <i>0.0007</i>	8
uk91	0.3381 <i>0.0053</i>	18	0.2612 <i>0.0029</i>	18	0.1706 <i>0.0025</i>	19
us91	0.3394 <i>0.0019</i>	19	0.2631 <i>0.0012</i>	19	0.1548 <i>0.0006</i>	9
ru92	0.4017 <i>0.0066</i>	20	0.2962 <i>0.0036</i>	20	0.1778 <i>0.0030</i>	20
mx89	0.4909 <i>0.0055</i>	21	0.3452 <i>0.0033</i>	21	0.2132 <i>0.0033</i>	21

Table 3: Ranking from LIS' Wave 4

LIS Country	Index (StdDev)	Ranking	Index (StdDev)	Ranking	Index (StdDev)	Ranking
$\alpha =$	0		0.25		1	
sw95	0.2218 <i>0.0019</i>	1	0.1848 <i>0.0012</i>	1	0.1496 <i>0.0008</i>	2
fi95	0.2257 <i>0.0028</i>	2	0.1890 <i>0.0016</i>	2	0.1508 <i>0.0012</i>	6
lx94	0.2353 <i>0.0043</i>	3	0.1984 <i>0.0029</i>	4	0.1553 <i>0.0019</i>	10
nw95	0.2403 <i>0.0049</i>	4	0.1972 <i>0.0029</i>	3	0.1521 <i>0.0024</i>	7
dk95	0.2532 <i>0.0026</i>	5	0.2077 <i>0.0015</i>	5	0.1503 <i>0.0011</i>	4
be97	0.2544 <i>0.0029</i>	6	0.2097 <i>0.0018</i>	6	0.1502 <i>0.0010</i>	3
nl94	0.2558 <i>0.0029</i>	7	0.2101 <i>0.0018</i>	7	0.1491 <i>0.0010</i>	1
cz96	0.2589 <i>0.0017</i>	8	0.2109 <i>0.0010</i>	8	0.1618 <i>0.0008</i>	13
ge94	0.2649 <i>0.0048</i>	9	0.2137 <i>0.0030</i>	9	0.1546 <i>0.0022</i>	8
rc95	0.2781 <i>0.0021</i>	10	0.2238 <i>0.0013</i>	10	0.1613 <i>0.0010</i>	12
cn94	0.2859 <i>0.0011</i>	11	0.2296 <i>0.0007</i>	12	0.1503 <i>0.0003</i>	5
fr94	0.2897 <i>0.0031</i>	12	0.2287 <i>0.0018</i>	11	0.1631 <i>0.0014</i>	14
as94	0.3078 <i>0.0028</i>	13	0.2442 <i>0.0017</i>	14	0.1549 <i>0.0011</i>	9
pl95	0.3108 <i>0.0024</i>	14	0.2393 <i>0.0014</i>	13	0.1641 <i>0.0011</i>	16
hu94	0.3248 <i>0.0081</i>	15	0.2491 <i>0.0048</i>	15	0.1683 <i>0.0037</i>	18
is97	0.3371 <i>0.0044</i>	16	0.2605 <i>0.0025</i>	17	0.1657 <i>0.0019</i>	17
it95	0.3406 <i>0.0037</i>	17	0.2604 <i>0.0021</i>	16	0.1639 <i>0.0015</i>	15
uk95	0.3429 <i>0.0041</i>	18	0.2630 <i>0.0023</i>	18	0.1732 <i>0.0020</i>	19
us94	0.3622 <i>0.0010</i>	19	0.2755 <i>0.0006</i>	19	0.1602 <i>0.0004</i>	11
ru95	0.4497 <i>0.0061</i>	20	0.3230 <i>0.0036</i>	20	0.1858 <i>0.0029</i>	20
mx96	0.4953 <i>0.0046</i>	21	0.3477 <i>0.0028</i>	21	0.2192 <i>0.0029</i>	21

Figure 6: The square root of P_1 's MSE for different bandwidths
Normal distribution, $\sigma = 1$

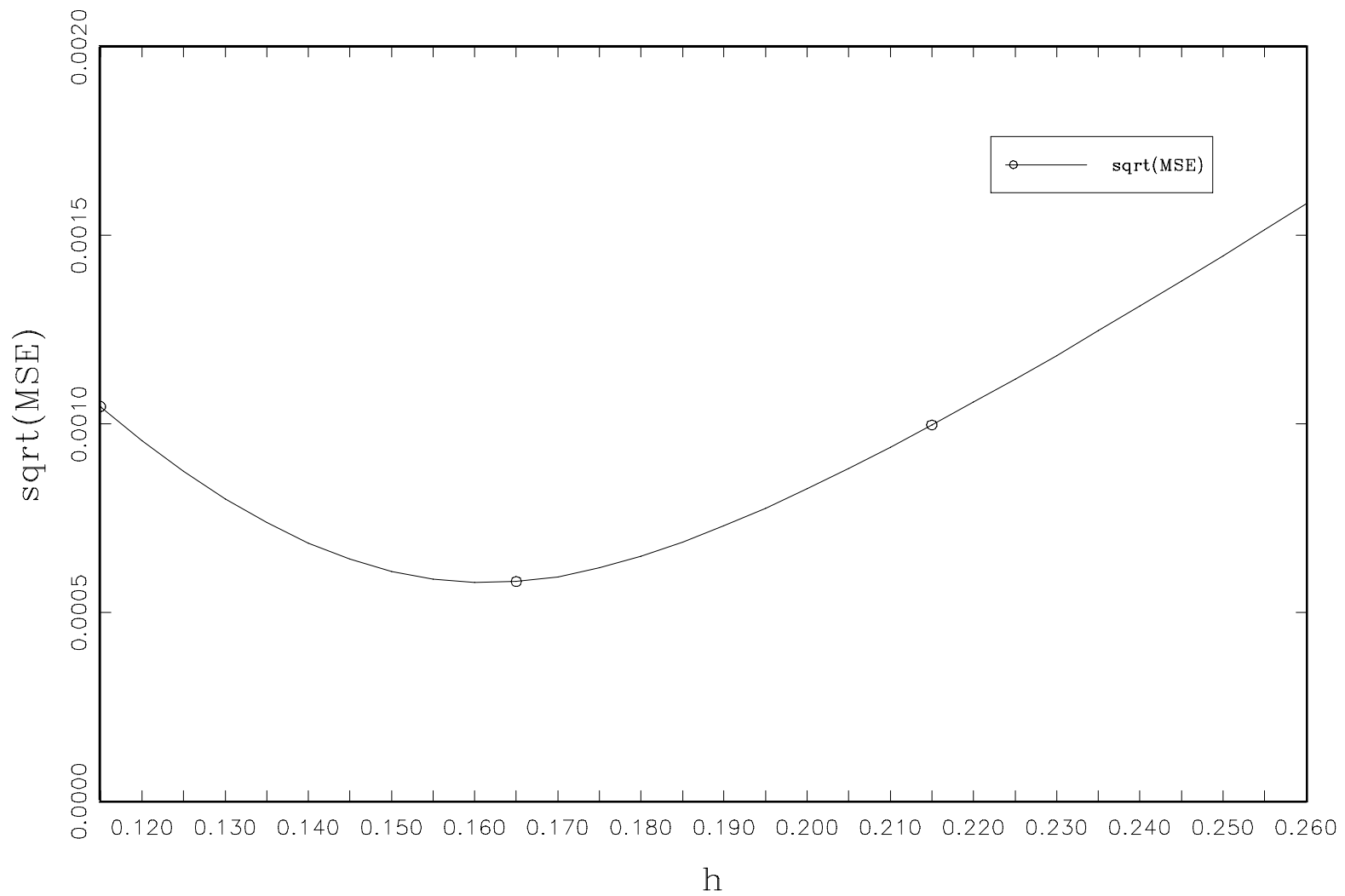


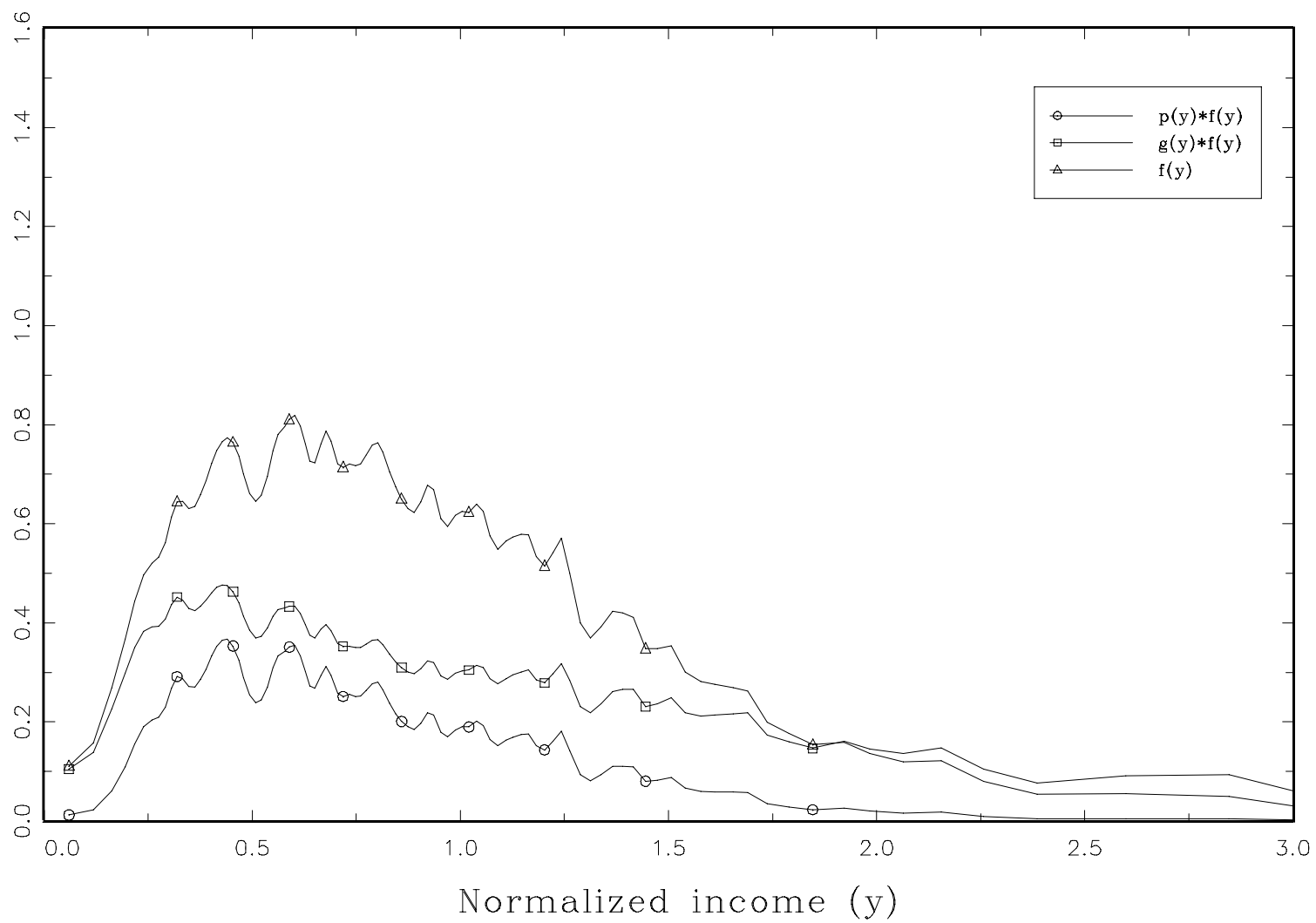
Figure 7: Polarization ($\alpha = 1$) and Inequality, United States 1991

Figure 8: Polarization ($\alpha = 1$) and Inequality, Czech Republic 1992

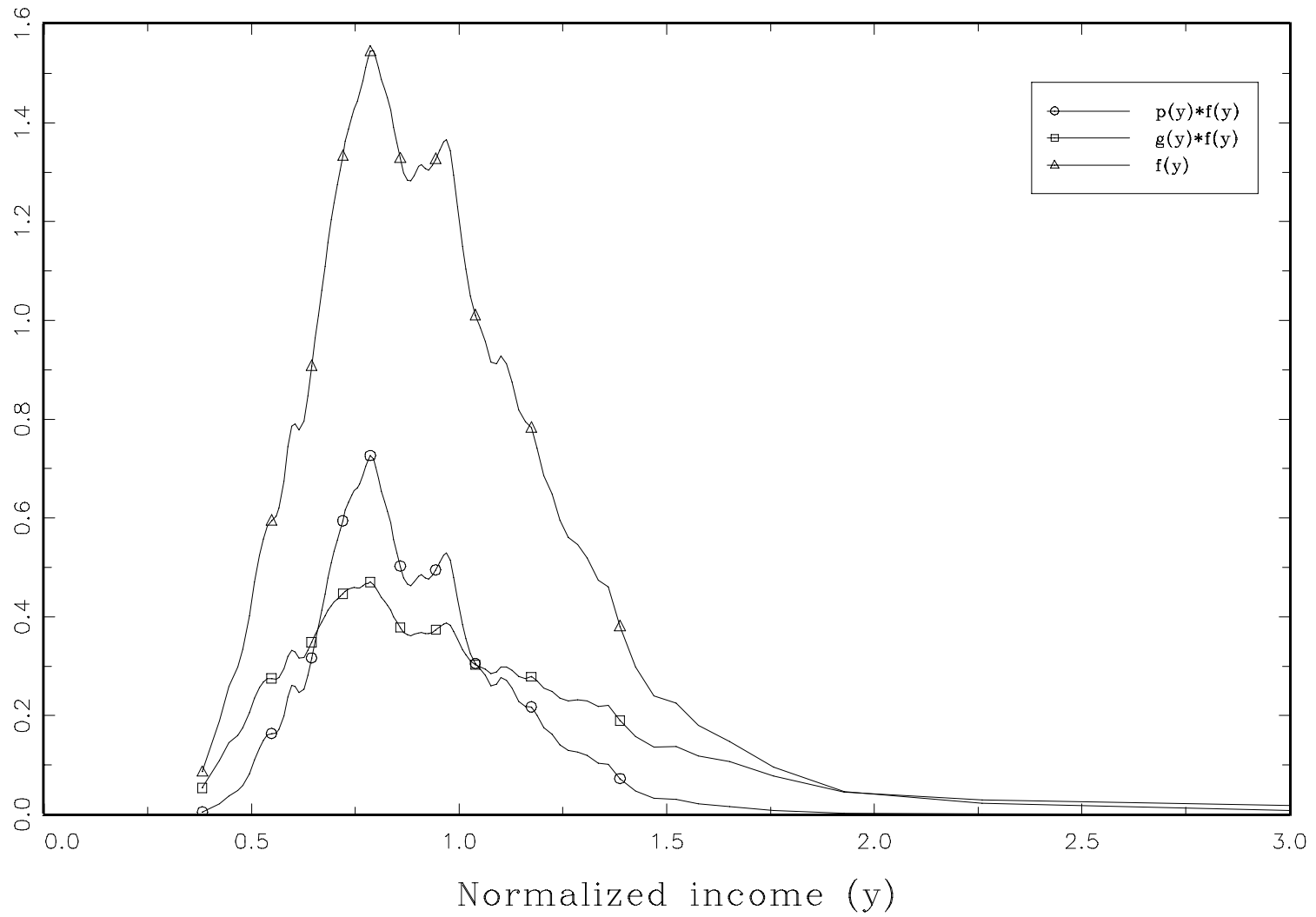


Figure 9: Differences in Polarization ($\alpha = 1$) and Inequality: Czech Republic 1992 and United States 1991

