Can the welfare of society be related to the degree of income polarization?*

JUAN GABRIEL RODRÍGUEZ †

Universidad Rey Juan Carlos de Madrid (Spain)

ABSTRACT

Measures of income polarization typically do not satisfy second-order stochastic dominance. Furthermore, in a polarized society, there might be no way of aggregating the opposing interests to achieve a social valuation of the alternative income distributions. In this paper, we obtain a class of social welfare functions that result from the trade-off between efficiency and income polarization. For this result, we make use of utility functions that depend on own income and the degree of personal complaints. This result is found for the Esteban and Ray (1994) and Duclos *et al.* (2004) polarization indices.

Key Words: welfare, income polarization, utility function.

JEL Classification: D39, D63, I31.

---

* I am grateful for helpful comments from J. M. Esteban and J.-Y. Duclos. This paper has benefited from the support of the Spanish Ministry of Education and Science [Project #SEJ2006-15172/ECON]. The usual disclaimer applies.

† **Address for correspondence**: Departamento de Economía Aplicada I, Universidad Rey Juan Carlos de Madrid, Campus de Vicálvaro, 28032 Madrid (Spain). E-mail: juangabriel.rodriguez@urjc.es.

1. Introduction

It is well known that one can express social welfare as resulting from mean income and a measure of income inequality. On this view, social welfare is the result of a trade-off between efficiency and equity. For decades, inequality has been the summary concept with which the distributional effects of changes in the economic environment have been evaluated. Nowadays, however, many economists rely on an alternative summary concept of polarization.[1] It is argued that polarization is a more appropriate criterion for explaining social conflict (see, for example, Esteban and Ray, 1999 and Reynal-Querol, 2002). Accordingly, one could ask: can the welfare of society be related to the degree of polarization? Does a trade-off between efficiency and social conflict exist? The answer to this question is by no means straightforward. Namely, there might be no way of aggregating the opposing interests to achieve a social valuation of the alternative income distributions in a polarized society. In fact, it might be that polarization and welfare cannot meaningfully be mixed together.

In this paper, we study this issue for income polarization based on the identification–alienation structure. We find that a social welfare function can result from the trade-off between efficiency and income polarization. In particular, we obtain a social welfare function for the polarization measures of Esteban and Ray (1994) and Duclos *et al.* (2004). To achieve these welfare functions, we make use of meaningful utility functions that depend not only on own income but also on the degree of personal complaints.

---

[1] See, for example, Foster and Wolfson (1992), Esteban and Ray (1994), Wolfson (1994), Wang and Tsui (2000), Gradín (2000), Zhang and Kanbur (2001), D'Ambrosio (2001), Chakravarty and Majumder (2001), Rodríguez and Salas (2003), Duclos *et al.* (2004), Duclos and Échevin (2005), Lasso de la Vega and Urrutia (2006) and Esteban *et al.* (2007).

The main contribution of the paper is not the proposed class of social welfare functions (arguably arbitrary) but the result that proves that a reasonable social welfare function that represents an hypothetical conflict between efficiency and income polarization can exist.

In the next section, we introduce the Esteban and Ray (1994) polarization index. Then we define some general restrictions that the social welfare function should satisfy and present our results. In Section 3, we apply our approach to the continuous case; that is, to the Duclos *et al.* (2004) polarization index. Finally, in Section 4, we conclude the paper with some remarks.

2. The discrete case

2.1. The Esteban and Ray (1994) polarization framework

Let $(\eta, x) = (\eta_1, ..., \eta_n; x_1, ..., x_n)$ be a distribution for any positive integer n if $x \in \mathbb{R}^n$, $x_i \neq x_j$, $\forall\ i, j$ and $\eta > 0$ where x is income. Total population is $\sum_{i=1}^{n} \eta_i$, and the mean income is μ.

Esteban and Ray (1994) (ER henceforth) assume that each individual is subject to two forces: *identification* with members considered to belong to the same group, and *alienation* from those considered to belong to other groups. Effective antagonism increases in identification and alienation in such a way that increased intragroup identification reinforces the effect of alienation. Polarization represents total effective antagonism. Accordingly, the ER polarization index is as follows:

$$P_{\alpha}^{ER} = C\sum_{i=1}^{n}\sum_{j=1}^{n}\eta_i\,\eta_j\eta_i^{\alpha}\left|x_i - x_j\right| \tag{1}$$

for some constants $C > 0$ and $\alpha \in [1,1.6]$ that represents the importance of group identification. The alienation term is $|x_i - x_j|$, and the identification term is η_i^{α}. We replace the population weights by the population frequencies by assuming that $C = \left(\sum_{i=1}^{n}\eta_i\right)^{-(2+\alpha)}$. In this case, expression (1) can be rewritten as follows:

$$P_{\alpha}^{ER} = \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i^{1+\alpha}\pi_j\left|x_i - x_j\right| \tag{2}$$

where π_i is the percentage of population of group i. Polarization in society is, therefore, the sum of all the effective antagonisms. The additive postulate is justified (following Harsanyi, 1953, among others) in terms of an impartial individual who might use the expected value of his or her effective antagonism to judge polarization in society.

Now, let W be a social welfare function defined on utilities that are represented by the function U. We want to know what restrictions have to be imposed on W and U so that:

$$W\left(U_1(x), U_2(x), \ldots, U_n(x)\right) = V\left(\mu, P_{\alpha}^{ER}\right) \tag{3}$$

for some function $V : S \subseteq \mathbb{R}_+^2 \to \mathbb{R}$, where $\frac{\partial V}{\partial \mu} > 0$ and $\frac{\partial V}{\partial P_{\alpha}^{ER}} < 0$. Expression (3) points out a trade-off between income polarization and efficiency. However, as said above, it might be that polarization and welfare cannot meaningfully be mixed together.

To study this issue, we initially impose the following set of restrictions:

1) $U_1(\cdot) = U_2(\cdot) = ... = U_n(\cdot) = U(\cdot)$, utility representations are identical for symmetry of the social welfare function.

2) U is not an individualistic function because people care not only about their own income but also about the income distribution they inhabit. Initial configurations of the whole distribution are relevant for the measurement of polarization. Accordingly, preferences cannot be individualistic if they are to take into account the whole distribution; that is, the utility function should be $U(x_i, x)$ for all $i = 1,.., n$.

3) Social welfare is an additively separable function; that is, $W(x) = \sum_{i=1}^{n} U(x_i, x) \cdot \pi_i$. The additive postulate is a common practice in polarization measurement (see, for example, Esteban and Ray, 1994 and Duclos *et al.*, 2004). To conform with this practice, we impose the requirement that the social welfare function be additive. In this way, impartial observers who evaluate overall polarization according to the expected value would apply the same criterion to welfare in society.

4) The social welfare function $W(x)$ is not Schur-concave. A function g: $\mathbb{R}^n \rightarrow \mathbb{R}$ is said to be Schur-concave if *Lorenz domination* of distribution x over distribution y implies that $g(x) \geq g(y)$. Thus, all Schur-concave functions satisfy the principle of progressive transfers. However, this principle is not fulfilled by income polarization indexes. Therefore, it cannot be that the social welfare function is Schur-concave.

Once we assume these restrictions, the natural question to ask is whether there exists any additively separable function W that satisfies (3). A negative answer would make a

plausible case for arguing that income polarization and welfare cannot meaningfully be mixed together.

2.2. Results

We start with a definition. Let $D(x_i;x_j)$ be the relative deprivation felt by an individual with income x_i in relation to an individual with income x_j, as follows (see Runciman, 1966, Yitzhaki, 1979 and Hey and Lambert, 1980):

$$D(x_i;x_j) = x_j - x_i \text{ if } x_i < x_j,$$

$$D(x_i;x_j) = 0 \text{ if } x_i \geq x_j. \tag{4}$$

Then, the deprivation felt by an individual with income x_i is:

$$D(x_i) = \sum_{j=1}^{n} D(x_i;x_j)\pi_j = \sum_{j=i+1}^{n} (x_j - x_i)\pi_j. \tag{5}$$

Given this ingredient, the following result proposes an additively separable function W that satisfies (3).

THEOREM 1. Given the social welfare function $W(x) = \sum_{i=1}^{n} U(x_i, x)\cdot\pi_i$ for every income distribution x:

$$U(x_i, x) = (1-\theta_i)x_i + \theta_i\left[\mu - 2D(x_i)\right] \text{ and } \theta_i = \pi_i^{\alpha} \Leftrightarrow W(x) = \mu - P_{\alpha}^{ER}.$$

Proof: The $\Rightarrow$ part. When we substitute equation (5) in the utility function definition, we obtain the following expression:

$$U(x_i, x) = (1-\theta_i)x_i + \theta_i \left[\mu - 2\sum_{j=i+1}^{n}(x_j - x_i)\pi_j \right].$$

We know that $\mu = \sum_{j=1}^{n} x_j \pi_j$, therefore we obtain the following:

$$\begin{aligned} U(x_i, x) &= x_i + \theta_i \left[\sum_{j=1}^{n} x_j \pi_j - 2\sum_{j=i+1}^{n} x_j \pi_j + 2\sum_{j=i+1}^{n} x_i \pi_j - x_i \right] \\ &= x_i + \theta_i \left[2\sum_{j=1}^{i} x_j \pi_j - \sum_{j=1}^{n} x_j \pi_j + 2\sum_{j=i+1}^{n} x_i \pi_j - \sum_{j=1}^{n} x_i \pi_j \right] \\ &= x_i - \theta_i \left[\sum_{j=1}^{n} \left(x_i + x_j - 2\min\{x_i, x_j\} \right) \pi_j \right]. \end{aligned}$$

Because $|x_i - x_j| = x_i + x_j - 2\min\{x_i, x_j\}$, we have:

$$U(x_i, x) = x_i - \theta_i \sum_{j=1}^{n} |x_i - x_j| \pi_j .$$

If $W(x) = \sum_{i=1}^{n} U(x_i, x) \cdot \pi_i$ and $\theta_i = \pi_i^{\alpha}$, then:

$$W(x) = \sum_{i=1}^{n} \left(x_i - \pi_i^{\alpha} \sum_{j=1}^{n} |x_i - x_j| \pi_j \right) \pi_i .$$

We need only consider expression (2) to complete the proof of this part.

The $\Leftarrow$ part. If we consider the proof above, this part is straightforward.

Now we comment on this result. First, we discuss the utility function. According to Theorem 1, the utility function of an individual with income x_i is the following:

$$U(x_i, x) = (1-\theta_i)x_i + \theta_i\,[\mu - 2D(x_i)] \qquad 0 < \theta_i = \pi_i^{\alpha} < 1. \tag{6}$$

The individual cares not only about his or her own income but also about the distribution to which he or she belongs. Utilities are, therefore, not individualistic. In fact, personal utility is a convex combination of two arguments: own income and income distribution in terms of mean income and deprivation. To understand the second argument fully, we note the following:

$$\begin{aligned} \mu - 2D(x_i) &= \sum_{j=1}^{n} x_j \pi_j - \sum_{j=i+1}^{n} x_j \pi_j + \sum_{j=i+1}^{n} x_i \pi_j - D(x_i) \\ &= \sum_{j=1}^{n} \min\{x_j, x_i\}\pi_j - D(x_i). \end{aligned} \tag{7}$$

The second argument implies that utility for people with income x_i will increase if the distance in terms of income between this group and the rest of the population decreases. That is, individual utility for a given personal income x_i will decrease not only when individual deprivation but also the gap with lower incomes increase. Temkin (1993) characterized inequality in terms of personal claims or complaints caused by income differences. If we borrow from Temkin (1993) his definition of personal claims, we can assume that the argument $\mu - 2D(x_i)$ defines the magnitude of complaints for an individual with income x_i. Accordingly, individual utility is a convex combination of own income and personal complaints. People are willing to give up some material payoff to move in the direction of more equitable outcomes.

Most economists routinely assume that material self-interest is the sole motivation of people. However, this practice contrasts with a large body of evidence gathered by psychologists and experimental economists during the last two decades. In particular, this evidence indicates that a substantial percentage of the population is strongly motivated by other-regarding preferences and that concerns for the well-being of others cannot be ignored in social interactions. An influential literature on social preferences and inequality aversion has recently pointed out these matters (see Sobel, 2005, Fehr and Schmidt, 2006 and the references therein). A feature that most of the models in this literature share is that individuals dislike inequality. Bolton and Ockenfels (2000), for instance, assume agents' preferences are an increasing function of their own payoff and their relative payoff. Fehr and Schmidt (1999) introduce a model with a similar motivation but assume an individual utility specification under which an agent cares about his own monetary payoff and, in addition, would like to reduce the inequality in payoffs across all agents. It turns out that our proposal could also be justified on the grounds of this literature.

Utility is, therefore, a balance between individual income and claims, but what is this balance? In the most general case, this balance will depend on individual characteristics like income, size of his or her group, socioeconomic and cultural background, and so on. In this case, every individual will have a particular weight for his or her income and claims. On the contrary, the most restrictive case will consist of a society where all individuals balance their incomes and claims likewise. We focus on an intermediate case where individuals balance both arguments according to the size of their group. In this manner, every income group has a different weight factor θ_i.

We assume a positive relationship between the relative weight that a person gives to the distribution he or she inhabits and the size of his or her group. The mechanism behind this relationship could be the following: if a group is small its members will feel themselves like an island inside society. This feeling of being apart from society will induce them to be more individualistic; that is, to give more weight to their incomes. In the extreme, if the group *i* becomes a single person ($\pi_i \to 0$), the individual will care only about his or her income ($U(x_i, x) \to x_i$). On the contrary, if a group is big, its members will feel that they constitute a relevant part of society, which will induce them to be more sympathetic with society as a whole. Accordingly, people will care more about the distribution and less about themselves. In the extreme, if the group *i* converges to the whole society ($\pi_i \to 1$), people will care about the mean income ($U(x_i, x) \to \mu$). A possible example of this at the country level is the case of Switzerland and the European Union. The former is a small country with few inhabitants that prefers, in general, not to take part in worldwide affairs; the latter is a set of countries with a large number of citizens that strives to implement policies worldwide.

Interestingly, the result in Theorem 1 can be replicated for a different polarization measure if we assume that weights depend on not only group size but also individual income. A polarization measure with an identification term that is positively correlated with not only the size of a group but also its income is commented on (though its axiomatic foundation remains undone) by ER. In this manner, the authors represented the case where more income signifies more power to make individual antagonism effective. This measure is $P_\alpha^{IER} = \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i^\alpha x_i^\beta |x_i - x_j| \pi_i \pi_j$ where $\beta \geq 0$ represents the sensitivity to

individual income of feelings of identification. It is straightforward to see that the result in Theorem 1 is obtained for this polarization measure when $\theta_i = \pi_i^{\alpha} x_i^{\beta}$. In this case, an individual will care more about others in society the larger the group and its income.

Finally, an increment in the income accruing to a person always has a positive effect on that person's utility. In formal terms, we know from the proof above that:

$$U(x_i, x) = (1-\pi_i^{\alpha})x_i + \pi_i^{\alpha}\left[\sum_{j=1}^{n} x_j \pi_j - 2\sum_{j=i+1}^{n}(x_j - x_i)\pi_j\right]. \tag{8}$$

Therefore, the first derivative of the utility function is the following:

$$\frac{\partial U(x_i, x)}{\partial x_i} = (1-\pi_i^{\alpha}) + \pi_i^{1+\alpha} + 2\pi_i^{\alpha}\sum_{j=i+1}^{n}\pi_j\ . \tag{9}$$

An increase of income *i* has a positive effect on own utility via own income, $1-\pi_i^{\alpha}$, the mean income, $\pi_i^{1+\alpha}$, and own deprivation, $2\pi_i^{\alpha}\sum_{j=i+1}^{n}\pi_j$. Consequently, $\frac{\partial U(x_i, x)}{\partial x_i} > 0$ for all $i = 1,..,n$.

Other observations about Theorem 1 are in order.

First, the social welfare function *W* is Paretian. According to the Pareto principle, if there is an increase in income of one person in society, other things remaining equal, social welfare will increase. For this result, we need to prove that $\frac{\partial W(x)}{\partial x_k} = \sum_{i=1}^{n}\frac{\partial U(x_i, x)}{\partial x_k}\pi_i > 0$ for all $k = 1,..,n$.

From (8) we obtain the following derivatives:

$$\frac{\partial U(x_i, x)}{\partial x_k} = -\pi_i^{\alpha} \pi_k \qquad for \;\; i = 1, ..., k-1, \tag{10}$$

$$\frac{\partial U(x_i, x)}{\partial x_k} = \pi_i^{\alpha} \pi_k \qquad for \;\; i = k+1, ..., n. \tag{11}$$

On the one hand, expression (10) represents the effect of an increase in the income of individual *k* over the utility of individuals with lower income. This effect is negative because the increase in personal deprivation more than compensates for the increase in mean income. On the other hand, expression (11) represents the effect of an increase in the income of individual *k* over the utility of individuals with higher income. This effect is positive because mean income increases while personal deprivation remains unchanged.

Consequently, if $\frac{\partial W(x)}{\partial x_k} > 0$ for $k = n$, which is the worst possible case, the social welfare function *W* will be Paretian. We know from equations (9) and (10) that:

$$\frac{\partial W(x)}{\partial x_k} = \sum_{i=1}^{k-1} \left(-\pi_i^{\alpha} \pi_k \right) \pi_i + \left[(1 - \pi_k^{\alpha}) + \pi_k^{1+\alpha} \right] \pi_k \tag{12}$$

where *k* = *n*. Now, if we take into account that $1 = \sum_{i=1}^{n-1} \pi_i + \pi_n$, the result will be the following:

$$\frac{\partial W(x)}{\partial x_n} = \left[\sum_{i=1}^{n-1} \left(\pi_i - \pi_i^{1+\alpha} \right) + (\pi_n - \pi_n^{\alpha}) + \pi_n^{1+\alpha} \right] \pi_n . \tag{13}$$

However, $\frac{\partial W(x)}{\partial x_n} > 0$ because $\pi_i > \pi_i^{1+\alpha}$ and $\pi_n \geq \pi_n^{\alpha}$; therefore, the social welfare function is Paretian.

Second, the rate of substitution between polarization and efficiency at a constant welfare level is unity; that is, $\frac{\partial P_\alpha^{ER}}{\partial \mu} = 1$. This means that social welfare is equally sensitive to mean income and income polarization. Note that the rate of substitution between inequality and efficiency is $\frac{\partial G}{\partial \mu} = \frac{1-G}{\mu}$ under the Sen (1974) welfare function $W = \mu(1-G)$, where G is the Gini index of inequality. In this case, social welfare is more sensitive to mean income than income inequality.

Finally, consider the ER polarization index with asymmetric alienation (AER):[2]

$$P_\alpha^{AER} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \pi_i^{1+\alpha} \pi_j (x_j - x_i) \tag{14}$$

where the poor feel alienated from the rich but the rich do not feel alienated from the poor. In this case, if we adopt the utility function $U(x_i, x) = x_i - \pi_i^{\alpha} D(x_i)$, our result for social welfare and income polarization, $W(x) = \mu - P_\alpha^{AER}$, will again be obtained. Utility depends only on own income and the deprivation felt by the individual. In this respect, it is worth noting that Cabrales *et al.* (2007) have recently proposed a model for analyzing the earning structure in the labor market. In their case, workers, in addition to the utility they obtain from their own wage, experience disutility from the wage of fellow workers

[2] This index is commented on by ER, though its axiomatic foundation remains undone.

who enjoyed similar circumstances in the near past and have a higher wage than their own. Moreover, it can be proved that this welfare function is also Paretian.

An alternative way of formulating this problem is by using the Duclos *et al.* (2004) polarization index for continuous distributions.

3. The continuous case

3.1. The Duclos *et al.* (2004) polarization framework

Let $f(x)$ be a frequency density function for $x \in [a, b]$, where x is income and $[a, b]$ is a bounded interval that contains the support of the distribution. We assume that $f(x)$ is differentiable on the open interval (a, b) and $\mu = \int_a^b x f(x) dx$ is the mean income. Moreover, the distribution function is $F(x) = \int_a^x f(y) dy$. Consider the Duclos *et al.* (2004) polarization index (DER henceforth) defined as:

$$P_\alpha^{DER}(F) = C \int_a^b \int_a^b f(x)^{1+\alpha} f(y) |y - x| \, dy dx \tag{15}$$

for some $C > 0$.[3] The sensitivity parameter $\alpha \in [0.25, 1]$ represents the importance of identification. Again the additive postulate is assumed; that is, polarization is the sum of all antagonisms.

Then, we may ask whether there exists a social welfare function W that satisfies the following:

---

[3] Homogeneity of degree zero is achieved by assuming that $C = \mu^{\alpha-1}$. This is equivalent to normalizing incomes by their mean.

$$W(F) = V\left(\mu, P_{\alpha}^{DER}\right) \tag{16}$$

for some function $V : Z \subseteq \mathbb{R}_{+}^{2} \to \mathbb{R}$, where $\frac{\partial V}{\partial \mu} > 0$ and $\frac{\partial V}{\partial P_{\alpha}^{DER}} < 0$. In a polarized society, once again, there might be no way of aggregating the opposing interests and achieving a social valuation of the alternative income distributions.

Let us impose the restrictions specified in Section 2 in continuous terms. First, we assume that all individuals have the same utility function $U(\cdot)$ for symmetry of the social welfare function. Second, utility is not individualistic; that is, the utility of an individual with income x in distribution $F(x)$ is $U(x, F)$. Third, social welfare is an additive separable function; that is, $W(F) = \int_a^b U(x, F) f(x) dx$. Finally, $W(F)$ is not Schur-concave.

Given our problem and these restrictions, we may think that the application to this continuous framework of the approach in Section 2 is, *mutatis mutandis*, a straightforward way to answer the question above. However, the continuous formulation is not restricted to distributions that are prearranged in groups. We show below that dealing with this fact will require a restriction on the constant C.

3.2. Results

Let $D(x)$ be the relative deprivation felt by an individual with income x (see Runciman, 1966, Yitzhaki, 1979 and Hey and Lambert, 1980):

$$D(x) = \int_x^b (y - x) f(y) dy. \tag{17}$$

More people with higher income increases personal deprivation.

Moreover, let $U(x,F)$ be the utility function of an individual with income x, as follows:

$$U(x,F) = (1-\theta_x)x + \theta_x\left[\mu - 2D(x)\right] \qquad 0 < \theta_x < 1. \tag{18}$$

As in the discrete case, individuals care not only about their own income but also about the distribution to which they belong. Personal utility is a balance between individual income and claims.

Given these two ingredients, the following result proposes an additively separable function W that satisfies (16).

THEOREM 2. Given the social welfare function $W(F) = \int_a^b U(x,F)f(x)dx$ for every income distribution $F(\cdot)$:

$$U(x,F) = (1-\theta_x)x + \theta_x\left[\mu - 2D(x)\right] \text{ and } \theta_x = C \cdot f(x)^{\alpha} \Leftrightarrow W(F) = \mu - P_{\alpha}^{DER}(F).$$

Proof: The $\Rightarrow$ part. When we substitute equation (17) in the utility function definition, we obtain the following expression:

$$U(x,F) = x - \theta_x x + \theta_x\left[\mu - 2\int_x^b (y-x)f(y)dy\right].$$

We know that $\int_a^b f(y)dy = 1$; therefore, we can rewrite the above expression as follows:

$$U(x,F) = x - \theta_x\int_a^b xf(y)dy + \theta_x\left[2\left(\mu - \int_x^b yf(y)dy\right) + 2\int_x^b xf(y)dy - \mu\right].$$

Furthermore, $\mu = \int_a^b yf(y)dy$, so we derive the following:

$$\begin{aligned} U(x,F) &= x + \theta_x \left[2\int_a^x yf(y)dy + 2\int_x^b xf(y)dy - \int_a^b yf(y)dy - \int_a^b xf(y)dy \right] \\ &= x + \theta_x \left[2\int_a^b \min\{y,x\} f(y)dy - \int_a^b yf(y)dy - \int_a^b xf(y)dy \right] \\ &= x - \theta_x \int_a^b \left(y + x - 2\min\{y,x\} \right) f(y)dy. \end{aligned}$$

Because $|y-x| = y + x - 2\min\{y,x\}$, we have the following:

$$U(x,F) = x - \theta_x \int_a^b |y-x| f(y)dy.$$

If $W(F) = \int_a^b U(x,F) f(x)dx$ and $\theta_x = C \cdot f(x)^\alpha$, then:

$$W(F) = \int_a^b \left(x - C \cdot f(x)^\alpha \int_a^b |y-x| f(y)dy \right) f(x)dx.$$

We need only consider expression (15) to complete the proof.

The $\Leftarrow$ part. If we consider the proof above, this part is straightforward.

This result has two potential difficulties with the weight factor $\theta_x = C \cdot f(x)^\alpha$. First, the constant C has no upper bound, and the term $f(x)^\alpha$ might be larger than unity.[4] Consequently, the weight factor θ_x could exceed unity, so utility would negatively depend on own income. To avoid this problem, we set an upper bound for the constant C.

[4] Nevertheless, empirical studies based on real data do not usually find values of $f(x)$ above one (see, for example, Duclos and Lambert, 2000 and Jenkins and van Kerm, 2005). See also below.

If we assume that $0 < f(x) < \infty$ for all x, we can consider $C = \frac{1}{K}$, for some constant $K > m_1 = \sup_x f(x)^{\alpha}$.

Moreover, we know from (17) and (18) that the utility function is as follows:

$$U(x,F) = x + C \cdot f(x)^{\alpha} \left[\mu - x - 2\int_x^b yf(y)dy + 2x\int_x^b f(y)dy \right]. \tag{19}$$

We also know that $\int_x^b yf(y)dy = \mu - \int_a^x yf(y)dy$, so applying integration by parts to $\int_a^x yf(y)dy$ we obtain:

$$\begin{aligned} U(x,F) &= x + C \cdot f(x)^{\alpha} \left[\mu - x - 2\left(\mu - xF(x) + \int_a^x F(y)dy \right) + 2x(1 - F(x)) \right] \\ &= x + C \cdot f(x)^{\alpha} \left[x - \mu - 2\int_a^x F(y)dy \right]. \end{aligned} \tag{20}$$

Therefore, the first derivative of $U(x, F)$ with respect to x is the following:

$$\frac{\partial U(x,F)}{\partial x} = 1 - C \cdot f(x)^{\alpha} \left[2F(x) - 1 + \frac{\alpha f(x)'}{f(x)} \left(2\int_a^x F(y)dy + \mu - x \right) \right]. \tag{21}$$

The first derivative of the utility function with respect to own income might be negative; that is, an increment to own income might reduce personal utility. If we assume that $0 < f(x) < \infty$ and $-\infty < f(x)' < \infty$, a positive sign for this expression will be guaranteed

for any income x if and only if $C = \frac{1}{K}$, for some constant

$$K > m_2 = \sup_x \left\{ f(x)^{\alpha} \left[2F(x) - 1 + \frac{\alpha f(x)'}{f(x)} \left(2\int_a^x F(y)dy + \mu - x \right) \right] \right\}.$$

Therefore, to solve both problems simultaneously, we should assume that $C = \frac{1}{K}$ for some constant $K > \max\{m_1, m_2\}$. To see how demanding this restriction is, we estimated m_1 and m_2 for the United States (2005) using the information on household incomes in the Panel Study of Income Dynamics (PSID) dataset and assuming that $\alpha = 1$. The results of the estimation were $\hat{m}_1 \cong 1.1 \cdot 10^{-5}$ and $\hat{m}_2 \cong 1.4 \cdot 10^{-5}$. It seems that the constraint on the constant C is not very demanding because both terms are rather small.[5]

---

[5] We know that the cumulative distribution function must satisfy $F(x) = G(x^*)$ where $x^* = x/\mu$. In turn, we obtain the following by taking derivatives: $\mu f(x) = g(x^*)$. Therefore, Theorem 2 will be well defined for a scale-free (homogeneous of degree zero) version of the Duclos *et al.* (2004) polarization index if we assume that $C = \frac{\mu^{\alpha-1}}{K^*}$ where $K^* > \max\{m_1^*, m_2^*\}$ with $m_1^* = \sup_{x^*} g(x^*)^{\alpha}$ and $m_2^* = \sup_{x^*} \left\{ g(x^*)^{\alpha} \left[2G(x^*) - 1 + \frac{\alpha g(x^*)'}{g(x^*)} \left(2\int_{a^*}^{x^*} G(y^*)dy^* + 1 - x^* \right) \right] \right\}$. This is equivalent to applying the proposed methodology to incomes normalized by their mean. The estimated values of m_1^* and m_2^* for the United States (2005) were 0.67 and 0.90, respectively. Again, the constraint on the constant C does not seem to be very restrictive. A more detailed description of the empirical results can be obtained from the author on request.

Given this value for the constant C, the first derivative of $U(x, F)$ with respect to x is positive. In turn, the derivative of $U(y, F)$ with respect to x is zero for all $y \in [a, b]$ distinct from x. Therefore, the social welfare function W is Paretian. Moreover, the rate of substitution between polarization and efficiency at a constant welfare level is again unity.

Finally, if we consider the Duclos *et al.* (2004) polarization index with asymmetric alienation (ADER):

$$P_{\alpha}^{ADER}(F) = C\int_{a}^{b}\int_{x}^{b} f(x)^{1+\alpha} f(y)(y-x)dydx \tag{22}$$

and adopt the utility function $U(x,F) = x - \left(C \cdot f(x)^{\alpha}\right)D(x)$ for some constants $C = \frac{1}{K}$ and $K > \sup_{x}\left\{ \frac{f(x)'}{f(x)^{1-\alpha}}\left(\int_{a}^{x} F(y)dy + \mu - x\right)\right\}$, our result between social welfare and income polarization will be obtained once again; that is, $W(F) = \mu - P_{\alpha}^{ADER}$. Note that the definition of K is simpler in this case.

4. Concluding remarks

Measures of income polarization do not fulfill the principle of progressive transfers. As a consequence, the welfare of society might not be related to the degree of income polarization. We prove in this paper, however, that such a relationship can be established, that welfare in society may result from a trade-off between efficiency and income polarization. We obtain this result for both the ER and Duclos *et al.*, (2004) polarization indices. One of the consequences of this result is that policy makers and researchers who

have usually justified their analyses of policy on the trade-off between efficiency and income inequality can now also implement their analyses of such policy on the basis of the trade-off between efficiency and income polarization.

The exercise presented in this paper applies to income polarization. However, the generalization of our result to social polarization measurement is by no means obvious. For instance, the additive separable hypothesis for the social welfare function might not be a reasonable assumption for some relevant social characteristics other than income. It is clear that further research on this issue must be undertaken.

References

Bolton, G. and A. Ockenfels, 2000. Erc: A theory of equity, reciprocity and competition. *American Economic Review* 90, 166–193.

Cabrales, A., A. Calvó-Armengol and N. Pavoni, 2007. Social preferences, skill segregation and wage dynamics. *Review of Economic Studies* 75, 65–98.

Chakravarty, S.R., Majumder, A., 2001. Inequality, polarization and welfare: theory and applications. *Australian Economic Papers* 40, 1–13.

D'Ambrosio, C., 2001. Household characteristics and the distribution of income in Italy: an application of social distance measures. *Review of Income and Wealth* 47, 43–64.

Duclos, J.-Y., Esteban, J., Ray, D., 2004. Polarization: concepts, measurement, estimation. *Econometrica* 72, 1737–1772.

Duclos, J.-Y., Échevin, D., 2005. Bi-polarization comparisons. *Economic Letters* 87, 249–258.

Duclos, J.Y. and P. Lambert, 2000. A normative and statistical approach to measuring classical horizontal inequity. *The Canadian Journal of Economics* 33, 87–113.

Esteban, J., Ray, D., 1994. On the measurement of polarization. *Econometrica* 62, 819–851.

Esteban, J.M. and D. Ray, 1999. Conflict and distribution. *Journal of Economic Theory* 87, 379–415.

Esteban, J., Gradín C., Ray, D., 2007. An extension of a measure of polarization, with an application to the income distribution of five OECD countries. *Journal of Economic Inequality* 5, 1–19.

Fehr, E. and K. Schmidt, 1999. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114, 817–868.

Fehr, E. and K. Schmidt, 2006. The economics of fairness, reciprocity and altruism—Experimental evidence and new theories in *Handbook on the Economics of Giving, Reciprocity and Altruism* (Eds.: S. Kolm and J. M. Ythier). North-Holland, Elsevier.

Foster, J. E., Wolfson, M., 1992. Polarization and the decline of the middle class: Canada and the U.S. Mimeo, Vanderbilt University.

Gradín C., 2000. Polarization by sub-populations in Spain, 1973–91. *Review of Income and Wealth* 46, 457–474.

Harsanyi, J. C., 1953. Cardinality utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy* 61, 434–435.

Hey, J.D., Lambert, P.J., 1980. Relative deprivation and the Gini coefficient: comment. *Quarterly Journal of Economics* 95, 567–573.

Jenkins, S.P., van Kerm, P., 2005. Accounting for income distribution trends: a density function decomposition approach. *Journal of Economic Inequality* 3, 43–61.

Lasso de la Vega, M.C., Urrutia, A.M., 2006. An alternative formulation of the Esteban–Gradín–Ray extended measure of polarization. *Journal of Income Distribution* 15, 10–42.

Reynal-Querol, M., 2002. Ethnicity, political systems, and civil wars. *Journal of Conflict Resolution* 46, 29–54.

Rodríguez, J.G., Salas, R., 2003. Extended bi-polarization and inequality measures. *Research on Economic Inequality* 9, 69–83.

Runciman, W.G., 1966. *Relative deprivation and social justice.* London: Routledge and Kegan Paul/Penguin Books.

Sen, A.K., 1974. Informational bases of alternative welfare approaches: aggregation of income distribution. *Journal of Public Economics* 3, 387–403.

Sobel, J., 2005. Interdependent preferences and reciprocity. *Journal of Economic Literature* 43, 392–436.

Temkin, L.S., 1993, *Inequality*. Oxford: Oxford University Press.

Wang, Y.Q., 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, 353–358.

Yitzhaki, S., 1979. Relative deprivation and the Gini coefficient. *Quarterly Journal of Economics* 93, 321–324.

Zhang, X., Kanbur, R., 2001. What difference do polarization measures make? An application to China. *Journal of Development Studies* 37, 85–98.