KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and
Applied Economics

Department of Economics

Weighting with Individuals, Equivalent Individuals or not Weighting at all. Does it Matter Empirically?

by

André DECOSTER
Erwin OOGHE

Public Economics

Center for Economic Studies
Discussions Paper Series (DPS) 02.15
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

October 2002

DISCUSSION
PAPER

Weighting with individuals, equivalent individuals or not weighting at all. Does it matter empirically?

André Decoster and Erwin Ooghe*
CES, KULeuven

May 26, 2002

1. Introduction

When income units are homogeneous in non-income characteristics, there exist many tools to make welfare comparisons between income distributions and the properties of these tools are well-understood (Lambert, 2001 for an overview). Unfortunately, these tools are too restrictive to make reasonable comparisons in practice, because:

> *At the heart of any distributional analysis, there is the problem of allowing for differences in people's non-income characteristics.* (Cowell and Mercader-Prats, 1999)

Therefore, practitioners have adapted standard tools in a simple way: first, convert the heterogeneous distribution of income and non-income characteristics into an equivalent distribution of living standards and second, apply the preferred standard tool to the distribution of living standards, either unweighted, or weighted by the number of individuals. The essence of the living standard function is thus to compress both income and non-income characteristics into a single comparable measure of well-being. This compression is often performed by dividing income by an equivalence scale factor to obtain equivalent incomes.

Glewwe (1991) suggests that transfers from poor to rich could decrease inequality. Both Ebert (1997) and Shorrocks (1995) analyze this problem formally and show that, under some standard assumptions, the problem essentially boils down to an incompatibility between Pareto indifference[1] – which requires indifference between two distributions with equal living standards for all individuals – and a *between type* Pigou-Dalton (BTPD) transfer principle – which socially prefers mean-preserving income

*CES, Naamsestraat, 69, B-3000 Leuven, Belgium. e-mail to andré.decoster@econ.kuleuven.ac.be or erwin.ooghe@econ.kuleuven.ac.be. This paper forms part of the research programme of the TMR network **Living Standards, Inequality and Taxation** [Contract No. ERBFMRXCT 980248] of the European Communities whose financial support is gratefully acknowledged. We are grateful to Christian Valenduc from the Belgian Ministery of Finance for making the microdata available to us, and for running the microsimulation program SIRe, to John Bishop, Peter Lambert, Erik Schokkaert and Fréderic Vermeulen for comments on an earlier version of the paper, and to Guy Van Camp for assistance with the computer programs. The authors alone are responsible for any errors.

[1] Ebert (1997) considers the stronger anonymity axiom defined over living standards.

transfers which make living standards more equal. The incompatibility can also be expressed in a different and more interesting way. When using equivalence scales to accomodate for differences in needs, the cited authors show that weighting income units by the (normalized) number of individuals is the only procedure satisfying Pareto indifference, whereas weighting income units by the (normalized) number of equivalent individuals is the only one satisfying the BTPD transfer principle. The ad hoc two step procedure of practitioners, as described above, satisfies Pareto indifference if one applies weighting by individuals, but not the BTPD transfer principle and the latter is precisely a more general way to translate Glewwe's message.

For heterogeneous welfare comparisons, Shorrocks (1995) prefers Pareto indifference and thus weighting with the (normalized) number of individuals, whereas Ebert (1997) seems to favour the BTPD transfer principle and weighting with the (normalized) equivalent number of individuals. Although it is essentially a choice between principles, a choice is not easily envisaged. We give two examples to make the difficulty clear and we therefore introduce some notation. We refer to income units by the index $i = 1, ..., n$ and y_i, n_i and m_i refer respectively to the income, the number of individuals and the equivalence scale (or the number of equivalent individuals) of income unit i. We assume a continuous, additively separable, strictly increasing and concave social welfare function defined over living standards. Following Pareto indifference, social welfare equals

$$W^A = \sum_i \frac{n_i}{n_\Sigma} V\left(\frac{y_i}{m_i}\right) \tag{1.1}$$

with $n_\Sigma = \sum_i n_i$ and the valuation function V is strictly increasing and concave. Following the BTPD transfer principle, social welfare equals

$$W^B = \sum_i \frac{m_i}{m_\Sigma} V\left(\frac{y_i}{m_i}\right) \tag{1.2}$$

with $m_\Sigma = \sum_i m_i$.

Consider the following simple example. Suppose two societies, which consist of singles (s) and couples (c). Singles are the reference type (equivalence scale equal to 1) and the equivalence scale of a couple equals 1.5. Society 1 contains a single and a couple, while society 2 has three singles. One single in both societies has an income – and living standard – equal to y. The couple in society 1 has a household income of 6 units and the living standard of each individual in this household equals 4. The other singles in the second society have an income equal to 4. The situation can be summarized in Example 1.

Example 1. The Pareto indifference principle

society 1				society 2			
household	s	c		household	s	s	s
income	y	6	⟷	income	y	4	4
equivalence scale	1	1.5		equivalence scale	1	1	1
living standard	y	4		living standard	y	4	4

Pareto indifference requires to rank both societies equal in terms of social welfare, since all individuals have the same living standard, irrespective of the value of y. Indeed, using W^A always leads to indifference since

$$\frac{1}{3}V(y) + \frac{2}{3}V(4) = \frac{1}{3}V(y) + \frac{1}{3}V(4) + \frac{1}{3}V(4)$$

Using W^B however, we will obtain indifference if and only if

$$\frac{1}{2.5}V(y) + \frac{1.5}{2.5}V(4) = \frac{1}{3}V(y) + \frac{1}{3}V(4) + \frac{1}{3}V(4)$$

which is only possible for all V if $y = 4$. In all other cases, we will have a strict preference for one of both societies, a rather counter-intuitive conclusion. Therefore, one might think Pareto indifference is indispensable for heterogeneous welfare comparisons.

Let us now turn to the between type Pigou-Dalton transfer principle. Consider again society 1 in example 1, but the incomes of the single and the couple equal respectively 5 and 15. Consider a money transfer of three units from the couple to the single. The situation of society 1 before and after the transfer can be summarized in Example 2.

Example 2. The between type Pigou-Dalton transfer principle

society 1 before transfer				society 1 after transfer		
household	s	c		household	s	c
income	5	15	$\longleftrightarrow$	income	8	12
equivalence scale	1	1.5		equivalence scale	1	1.5
living standard	5	10		living standard	8	8

A money transfer which makes living standards more equal always has to raise social welfare according to the BTPD transfer principle. Indeed, using W^B, social welfare before the transfer equals

$$\frac{1}{2.5}V(5) + \frac{1.5}{2.5}V(10)$$

which is, by definition of strict concavity, always smaller than

$$V\left(\frac{1}{2.5}5 + \frac{1.5}{2.5}10\right) = V(8)$$

which is precisely social welfare after the transfer. Using W^A however, social welfare before the transfer equals

$$\frac{1}{3}V(5) + \frac{2}{3}V(10)$$

which is not necessarily smaller compared to social welfare after the transfer, $V(8)$. More precisely, it will depend on the concavity of the utility function. To make things clear, consider a (concave) iso-elastic utility function

$$V: \begin{array}{ll} x \to \frac{1}{1-\varepsilon} x^{1-\varepsilon}, & \varepsilon \neq 1 \\ x \to \ln x, & \varepsilon = 1 \end{array}$$

with $\varepsilon > 0$ measuring the degree of concavity. Now, choosing ε extremely small (let $\varepsilon \to 0$ and W^A comes arbitrarily close to the average living standard) we will prefer the society before the transfer to the society after the transfer. Choosing ε extremely large (let $\varepsilon \to \infty$ and W^A comes arbitrarily close to the (lexicographical) minimum living standard) we will prefer the society after the transfer to the society before the transfer. To conclude, if one thinks the BTPD transfer principle is always interesting, one might be inclined to use W^B rather than W^A, since the latter only satisfies it when the degree of concavity is sufficiently large.[2]

Both examples indicate that a choice between the two principles – and thus also between the corresponding weighting principles – is not easy. But, if the difference between both normative properties does not translate into a different comparison in most empirical applications, it could be rather futile. Accordingly, the question we raise in our paper is simple. Do we observe large differences between the two alternative weighting methods empirically? We certainly do not claim that a normative choice can ultimately be reduced to an empirical question and that this question can be solved by looking at one particular case. But at least we hope to shed light on this question which, as far as we know, has not been treated elsewhere before.

Our paper is organised as follows. The next section introduces some notation and presents our methodology in a formal way. Section 3 describes the data and section 4 shows our main results. A final section summarizes the discussion as well as our main findings.

2. A framework for the social planner

A social planner is facing two distinct questions. The first one is how to measure the well-being of individuals given their household income and household composition. The other is how to aggregate these well-beings to obtain a measure of social welfare. We will first present the necessary notation before answering both questions.

Income units are fully characterized by their income $y \in \mathrm{IR}_+$ and non-income characteristics and we assume that we can divide the population in a finite number of types $k = 1, ..., K$ based on these non-income characteristics. We will consider differences in household size n_k and we allow for a different treatment of adults and children. More precisely, the number of adults and children in type k households equal respectively n_k^a and n_k^c, and thus for all k we need $n_k = n_k^a + n_k^c$.

The ultimate goal of a social planner is to make heterogeneous welfare comparisons, i.e. welfare comparisons between heterogeneous distributions of nominal household

[2] Notice that there exists a rule which (more than) satisfies both the BTPD transfer principle and Pareto indifference, to wit, the leximin rule, but this rule is excluded in Ebert (1997) and Shorrocks (1995), since both authors start from a continuous social welfare function.

incomes, F and G for instance. Both distributions are defined over a common finite income support and can be decomposed in subdistributions F_k (or G_k) and proportions p_k (or q_k) for each type k such that $F(y) = \sum_k p_k F_k(y)$ (or $G(y) = \sum_k q_k G_k(y)$) measures the proportion of households with an income lower or equal to y under distribution F (or G).

2.1. Measuring well-being

The well-being of individuals is measured simply by correcting household income y through an equivalence scale factor $m_k > 0$, or for all types k, living standard functions are simply defined as:

$$v_k : \mathrm{IR}_+ \to \mathrm{IR}_+ : y \to v_k(y) = \frac{y}{m_k} \tag{2.1}$$

We stress that $v_k(y)$ measures the living standard of *each* individual who lives in a household of type k with household income y. The equivalence scale m_k reflects "needs" in the following simple way. A household with a larger equivalence scale needs more income to reach the same living standard.

One simple way to correct incomes is to use a parametric equivalence scale (Banks and Johnson, 1994), defined as

$$m_k = \left(n_k^a + \eta n_k^c\right)^\theta, \ 0 \le \theta \le 1, 0 \le \eta \le 1 \tag{2.2}$$

η is the way in which children are converted in adults and we will choose $\eta = 0.5$ (Jenkins and Cowell, 1994). We can now define adult-equivalents as $\widetilde{n}_k = n_k^a + 0.5 n_k^c$. The parameter θ is the equivalence scale elasticity, measuring the percentage change in equivalence scale m for a (small) percentage change in the number of adult-equivalents. Although we will present our results for all θ, it is worth noting that reasonable values range between approximately 0.4 and 0.8 (Burkhauser *et al.*, 1996). We can rewrite equation (2.2) as:

$$m_k = \left(\widetilde{n}_k\right)^\theta, 0 \le \theta \le 1 \tag{2.3}$$

Using both equations (2.1) and (2.3), we get

$$v_k : \mathrm{IR}_+ \to \mathrm{IR}_+ : y \to v_k(y) = \frac{y}{\left(\widetilde{n}_k\right)^\theta} \tag{2.4}$$

Independent of a specific choice for θ, some interesting properties are satisfied. Notice that singles without children ($n^a = 1$, $n^c = 0$) are the reference group, since their equivalence scale always equals 1, for all θ. Furthermore, for extremely low (0) or extremely high (∞) household incomes, differences in household size do not imply differences in "needs". Finally, "needs" are bounded in several interesting ways. They are clearly increasing in adult-equivalents. In addition, larger households (in adult-equivalents) have scale advantages in "needs", since they need not less income than the reference type, but also not more than $\widetilde{n}_k$ times the income of the reference type, to reach the same living standard.

2.2. Aggregating well-beings

We will consider two standard ways of aggregating well-beings into a measure of social welfare, more precisely, via the family of S-Gini social welfare functions (Donaldson and Weymark, 1981), and via a welfare dominance procedure (Shorrocks, 1983). In general, we can write social welfare of a distribution F in terms of living standards $x = \frac{y}{m_k}$ simply as:

$$W_F = \sum_k \int_0^\infty V(x, F^*)\, dF_k^*(x) \tag{2.5}$$

Social welfare consists of two parts: (i) the valuation of living standards x via a valuation function V – which might be non-individualistic by taking the weighted distribution F^* into account – and (ii) the actual aggregation of these valuations via weighted distributions F_k^*.

As should be clear from the introduction, choosing either individuals or equivalent individuals as the unit of analysis is a normative choice. The weighted distribution F^* can be generally defined as

$$F_k^* : x \to p_k^* F_k(m_k x) \tag{2.6}$$

For both weighting procedures, we have to choose the appropriate weighting vector $w = (w_1, ..., w_K)$ to obtain normalized proportions p_k^* equal to $\frac{p_k w_k}{\sum_k p_k w_k}$ for each type k. Following Shorrocks (1995), we could weight households by the number of individuals ($w_k = n_k$ for all k) and p_k^* represents the proportion of individuals belonging to type k. Following Ebert (1997), we can weight households by the equivalence scale factors ($w_k = m_k$ for all k) and p_k^* becomes the proportion of adult-equivalents. Finally, we also consider the possibility of not weighting at all ($w_k = 1$ for all k) since this convention has been often used in practice.[3] In this case p_k^* simply equals the proportion of households p_k. We stress that normatively speaking this scenario is not interesting since it satisfies neither Pareto indifference nor the BTPD transfer principle. Notice also that using equivalence scales as weights may coincide with not weighting at all, if $\theta = 0$.

Whereas both procedures aggregate in the same way, their valuation functions – containing the ethical view of the social planner – are very different. The S-Gini social welfare functions weight living standards by a factor which depends on the position of the income unit in the overall (weighted) distribution of living standards in the following way (see Donaldson and Weymark, 1983):

$$V : (x, F^*) \to \delta \left(1 - \sum_k F_k^*(x)\right)^{\delta - 1} x \tag{2.7}$$

Depending on the weighting scheme, the term between brackets $1 - \sum_k F_k^*(x)$ measures the proportion of individuals, equivalent individuals or households with a

[3] Bishop, Chow, Formby and Ho (1997, p.191) even refer to this difference of household or person weighting as one of the possible explanations for divergent empirical results in studies of changes in US tax progressivity.

living standard higher than x. Higher living standards get lower weights. In contrast with the standard Gini valuation function, the S-Ginis add a distributional parameter δ, $1 \leq \delta \leq \infty$, which defines the relative weight attached to lower living standards. Higher values for δ correspond with a relative higher concern for inequalities in living standards. The limiting cases correspond with utilitarianism ($\delta \to 1$) and the maximin social welfare function ($\delta \to \infty$), while the standard Gini social welfare function can be obtained when $\delta = 2$.

The valuation function in the welfare dominance procedure does not depend on the weighted distribution F^*:

$$V^\circ : (x, F^*) \to V^\circ (x, F^*) = V(x) \tag{2.8}$$

The twice differentiable valuation function contains the ethical view of the social planner and we will stick to the standard assumptions that social welfare increases when (1) living standards are higher ($V^{'} > 0$ everywhere) and (2) living standards are more equal ($V^{''} < 0$ everywhere). The family of valuation functions satisfying both of these conditions is denoted $\mathcal{V}$.

Let us briefly summarize. Either we use one of the S-Gini social welfare functions by substituting equations (2.7) into (2.5):

$$W_F = \sum_k \int_0^\infty \delta \left(1 - \sum_k F_k^*(x)\right)^{\delta-1} x dF_k^*(x), \ 1 \leq \delta \leq \infty \tag{2.9}$$

Or, we use the welfare dominance procedure which looks for unanimity among transformed "utilitarian" social welfare functions, obtained by substituting equation (2.8) into equation (2.10):

$$W_F = \sum_k \int_0^\infty V(x) \, dF_k^*(x), \ V \in \mathcal{V} \tag{2.10}$$

We want to analyze whether the choice of weighting scheme affects the results, more precisely the sign of $\Delta W = W_F - W_G$. However, whereas a particular S-Gini leads to a real number for ΔW, this is not the case for the welfare dominance procedure, since the latter leads to an infinity of values for ΔW, one for each valuation function belonging to $\mathcal{V}$. Fortunately, there exists a more practical procedure. We define the welfare dominance quasi-ordering $\succsim_{(\theta,w,\mathcal{V})}$ such that $F \succsim_{(\theta,w,\mathcal{V})} G$ means that F welfare dominates G, given a choice for the equivalence scale elasticity θ, the weighting vector w and the family of valuation functions $\mathcal{V}$. More specifically, F welfare dominates G if and only if ΔW is positive for all transformed utilitarian social welfare functions corresponding with the specific choice for $(\theta, w, \mathcal{V})$. The following lemma shows a classical equivalence result between welfare dominance $\succsim_{(\theta,w,\mathcal{V})}$ and generalized Lorenz dominance (GLD):

Lemma 1. $F \succsim_{(\theta,w,\mathcal{V})} G \Leftrightarrow$

$$\sum_k H_k(a) = \sum_k \int_0^a \left[F_k^*(x) - G_k^*(x)\right] dx \leq 0, \forall a \in I\!R_+ \tag{2.11}$$

3. The data

Using the microsimulation model SIRe of the Ministery of Finance, we calculate tax liabilities and disposable income before and after a recently decided personal income tax reform.[4] Although this reform will be implemented only gradually, we simulate the complete reform, assuming, however, no changes in labour supply decisions overall. Let us present briefly the main features of the tax reform.

First, the top marginal tax rates (52.5% and 55%) are abolished, while the middle tax brackets are widened considerably. Second, a refundable tax credit is introduced. Over a range of low labour incomes, this tax credit first gradually increases with income, stays constant over a range to decrease again gradually. Third, the discrimination between a married couple and a cohabitating couple is removed. On the one hand, the tax exemption for a married couple will be set at the level of two (possibly cohabitating) singles and taxes on non-labour incomes (income from financial assets or property) of married individuals will be levied separately. On the other hand, the marital quotient – the possibility of shifting taxable income between spouses – will be equally applicable for cohabitating singles.

The tax reform is not revenue neutral. We denote with $t^0(y)$ and $t^1(y)$ the tax liability for a given gross income y before and after the tax reform and $d^0(y) = y - t^0(y)$ and $d^1(y) = y - t^1(y)$ represent the corresponding disposable incomes. T^0 and T^1 represent aggregate tax liabilities and D^0 and D^1 aggregate disposable income. The percentage change in total taxes $\lambda = \frac{T^1 - T^0}{T^0}$ equals -11.8%. The percentage change in aggregate disposable income, denoted by $\rho = \frac{D^1 - D^0}{D^0}$, equals +3.9%. Given such a substantial tax reduction, a GL-dominance analysis will tend to favour the tax reform. Therefore we have constructed two revenue neutral benchmarks. First, we compare the tax reform with a reform – defended by social democrats – which would give all households the same percentage increase in disposable income. We call this benchmark the 'RP-neutral' one, since it holds residual progressivity constant (Lambert, 2001 p.220-221). We compare disposable incomes after the tax reform, $d^1(y)$, with RP-adjusted disposable incomes, $d^{RP}(y)$, with $d^{RP}(y) = (1+\rho)\, d^0(y)$. Second, we assess the tax reform relative to a benchmark – defended by liberal democrats – which would give all households the same percentage reduction in tax liabilities. In this case, liability progressivity is held constant, and we therefore refer to this benchmark as the 'LP-neutral' one. We now compare disposable incomes after the tax reform, $d^1(y)$, with LP-adjusted disposable incomes $d^{LP}(y)$, with $d^{LP}(y) = y - (1+\lambda)\, t^0(y)$. We stress that both RP- and LP-neutral benckmarks are calculated on the basis of incomes rather than living standards and, as a consequence, average living standards after taxes and after the benchmark reform are not necessarily the same.

4. Empirical results

We estimate the Generalized Lorenz curves and the S-Gini social welfare function on 9 ordinate points (we worked with deciles). An often neglected issue is that increasing

[4] To limit the influence of outliers we have trimmed the dataset with 1% on both upper and lower tail.

the number of intervals may seriously influence the dominance results, and/or the number of crossings (see Howes, 1996). Furthermore, no statistical inference has been performed for our results. Evidently this should be a topic for future research.

A social planner ultimately wants to know whether one of the heterogeneous income distributions is better in terms of social welfare than the other. In our case, he is interested whether the tax reform is better or worse compared to some benchmark tax reform. We can expect a social planner to prefer a robust comparison, so he will logically start with the most general procedure to compare both heterogeneous distributions, to wit the GLD procedure.

4.1. Welfare comparisons by means of the GLD procedure

Before analyzing the dominance results for the different weighting methods, we want to stress a crucial issue with respect to the interpretation of the results. When comparing the effect of two different weighting methods on dominance analysis, we can basically find three results which have to be distinguished cautiously. Suppose first that, for some values for θ, we find dominance in case of one weighting method and the opposite dominance for another weighting method. In this case, the weighting undeniably plays a crucial role. We can say that, for each of these peculiar values for θ, both weighting methods lead to contradictory policy conclusions for *all* valuation functions in $\mathcal{V}$.

The second possibility is that, for some values for θ, we find dominance in case of one weighting method and no dominance for another weighting method. Now we can only state that, for each of these values for θ, there exists *at least one* valuation function in $\mathcal{V}$ for which the policy recommendation contradicts. Finally, if we find, for some values of θ, that there is no dominance in case of two weighting methods, then we cannot say anything at all. At first sight, one could be inclined to conclude that both weighting methods lead to the same result, to wit, no dominance, but this, of course, is a too hasty conclusion. It might be the case that, for these values of θ, all valuation functions are contradictory but it might also be the case that none are actually contradictory. To conclude, the first case would give a very clear answer to the question whether weighting is empirically important, at least in case of our tax reform evaluation, while the other two cases give a very vague or no idea at all about the practical importance of weighting differently.

Figure 1 presents the dominance results for different values of θ when using the RP benchmark. The horizontal axis displays parameter θ from 0 to 1. The bars represent the results for the three different weighting schemes, shaded differently depending on the result obtained.

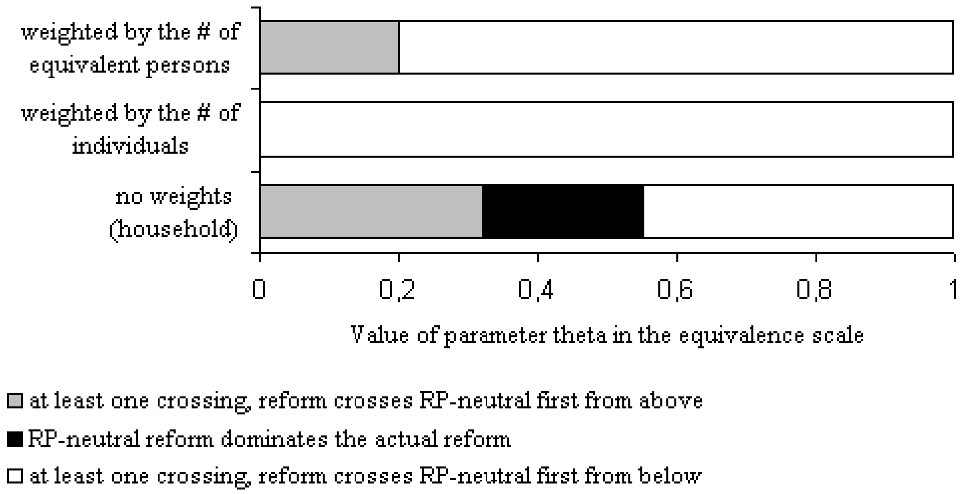

Figure 1: GLD comparison of the tax reform with the RP-neutral reform

With respect to the *results* themselves, we basically find no dominance, except when we use households as the unit of analysis. In the latter case, the RP-neutral tax reform dominates the tax reform over a certain range of θ-values (the black part of the bar). To put it differently, the tax reform is inferior to a tax reform which would have increased disposable incomes proportionately. Second, in case of no dominance there are two possibilities. Either the tax reform curve cuts the benchmark tax reform curve from above (the shaded part of the bar). This means that valuation functions with a high enough concern for inequality in living standards (a high degree of concavity) will prefer the tax reform to the benchmark tax reform. Or, the tax reform curve cuts the benchmark tax reform curve from below (no shading) which would, for the same type of valuation functions, lead to the reverse conclusion.

For reasonable values of θ, the tax reform curve cuts the benchmark curve first from below (for weighting by the number of individuals even over the whole θ-range). This means that for sufficiently concave valuation functions the RP-benchmark is preferred to the current tax reform. Only for low values of θ, the first crossing comes from above in the case of weighting by household or by equivalent individuals, indicating that for sufficiently inequality averse valuation functions, the tax reform might be welfare superior to the RP-benchmark.

With respect to the *weighting* issue, we cannot say much, unfortunately. As should be clear from the introduction, those practitioners who choose households as the unit of analysis should be aware that their choice is normatively doubtful, since none of both attractive principles are satisfied. In addition, this choice could be relevant empirically, or more precisely, for each value of θ within a range of approximately $[0.3, 0.55]$, there exists at least one valuation function in the set $\mathcal{V}$ such that switching to individuals or equivalent individuals as the unit of analysis would lead to the opposite conclusion (since the dominance gets lost). We will need a more fine-tuned approach if we want to give a more interesting meaning to "at least one".

We now turn to GL-dominance with respect to the LP-neutral tax reform, presented in Figure 2. We again start with a discussion of the result itself. In the case where the individual is chosen as the appropriate unit of analysis, we always find that the tax reform is better compared to the LP-neutral one (the black bar). This is consistent with the rather safe assumption that the tax reform is a compromise between the RP-neutral and the LP-neutral tax reform. In the case of using households, we never find dominance. But again, looking at the crossing of both curves, when using valuation functions which do care enough about inequalities in living standards, we obtain preference for the tax reform over an LP-neutral reform.

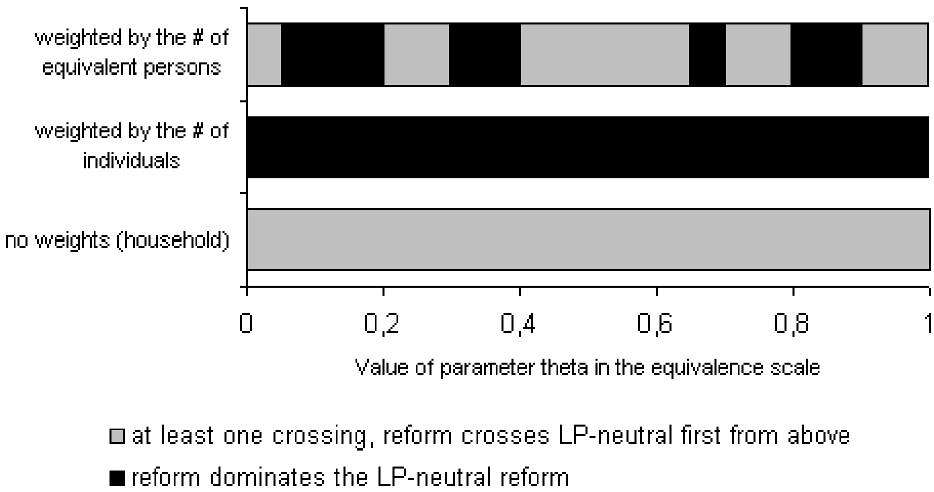

Figure 2: GLD comparison of the tax reform with the LP-neutral reform

Using equivalent individuals leads to very sensitive results. The result switches between no dominance and dominance. Figure 2 is based on a grid over θ-values in steps of 0.05. But a finer grid with a stepsize of, e.g. 0.001, only corroborates the sensitivity result. In many cases there was a switch between dominance and no dominance for each increment in θ.[5] If this extreme sensitivity is confirmed in other studies, and if it proves to be robust with respect to statistical tests, this might point to an unexpected drawback of this weighting procedure. Given the fact that there is no consensus about the right value for θ, using equivalent individuals as the unit of analysis in generalized Lorenz dominance could turn out to be a rather indecisive procedure if we want to obtain robust results with respect to the choice of equivalence

[5] For obvious reasons we could not display this extremely fanciful pattern in a figure.

scales. With respect to the effect of the different weighting methods, we cannot say much. For all values of θ, there exists at least one valuation function in $\mathcal{V}$ such that switching between individuals and households leads to a contradictory result. Since using equivalent individuals is, for all values of θ, different from exactly one of the other weighting methods, we get a similar, but unfortunately, minimal conclusion.

To summarize our generalized Lorenz dominance investigation, mainly two results are worth repeating. For the results, an instability arises when using adult-equivalents as the unit of analysis. For the weighting issue itself however, we are less successful. Up till now we have to content ourself with finding "at least one" valuation function for which weighting is crucial. If we want to answer the weighting question we need to find a more precise quantitive statement of "at least one" and we therefore turn to the S-Ginis in the next subsection.

4.2. Welfare comparisons by means of S-Gini social welfare functions

We start our analysis with applying the standard Gini social welfare function ($\delta = 2$ in equation (2.9)) to our data. Figure 3 illustrates our results, for different values of θ and for the different weighting methods. Only for large enough values of θ, we find that the three methods agree and prefer the actual reform to the RP-neutral one. But for lower values of θ there is a conflict between weighting by individuals on the one hand, and the two other procedures on the other hand. Individual weighting always concludes that the actual reform is welfare superior with respect to the RP-neutral one. But the other two methods (household and equivalent weighting) lead to the opposite conclusion. We do not present a figure for the comparison with the LP-neutral benchmark since all methods prefer the tax reform to the LP-neutral one for all values of θ. Of course, the particular choice for the distributional parameter δ is rather arbitrary and therefore we propose an S-Gini dominance procedure for all values of δ between 1 and 5 and the results w.r.t. the RP-neutral benchmark are presented in Figure 4.

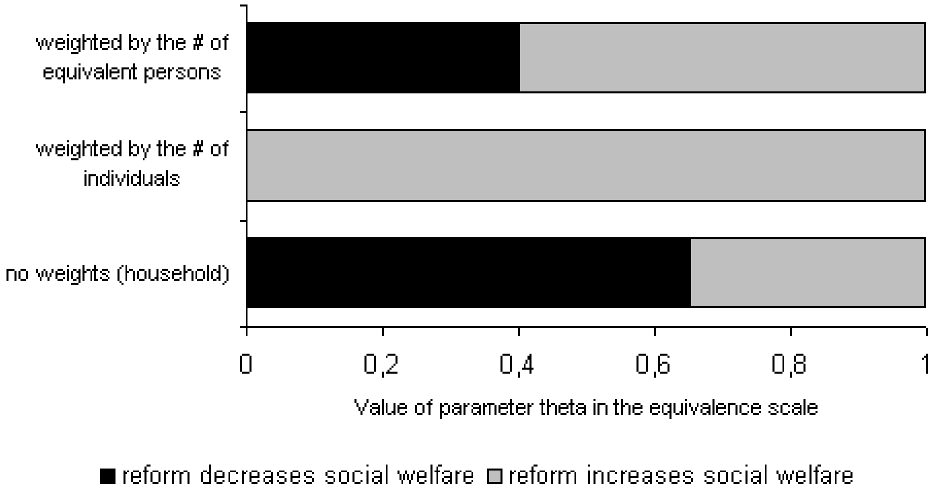

Figure 3: S-Gini social welfare comparison ($\delta = 2$) of the tax reform with the RP-neutral reform

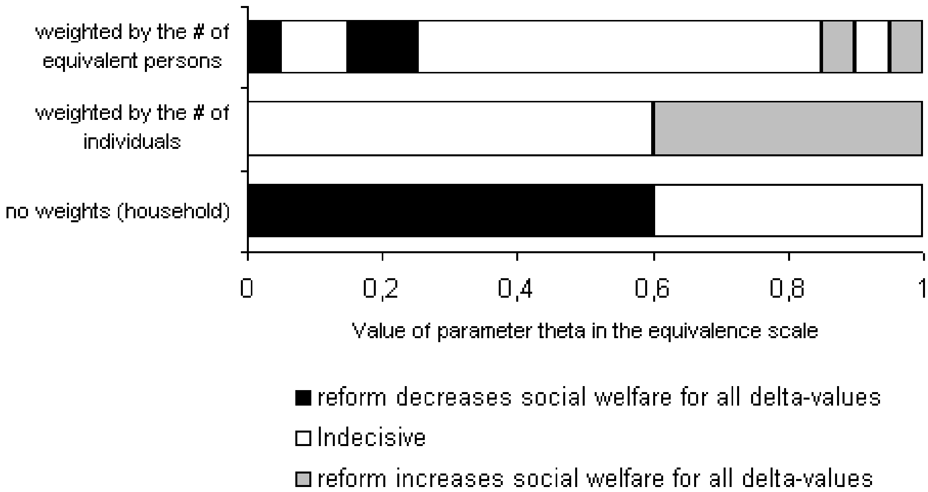

Figure 4: S-Gini social welfare comparison ($1 \leq \delta \leq 5$) of the tax reform with a RP-neutral reform

Since the S-Gini social welfare functions are Lorenz consistent, they can be considered to be a subset of the set of of valuation functions $\mathcal{V}$, for which we have inspected the GLD criterion. Hence it will not come as a surprise that we find more dominance results in figure 4 when compared with figure 1. The limited dominance result for household weighting in figure 1, has now been extended over the whole range of θ-values from 0 up to .6. Also for individual weighting, where no dominance was found in figure 1, we find a dominance result now. For θ-values from .6 to 1.0, the tax reform increases social welfare for the specified subset of S-Gini social welfare functions. In the light of our research question, the fact that this reversal of the dominance pattern between household and individual weighting only occurs at different values of θ, is but scant comfort for the practitioner. Within a reasonable range of θ-values, the dominance result does change. Moreover, for all values of θ, we know there exists at least one valuation function where the conclusion is reversed. Finally, looking at the procedure of equivalized weighting, the picture still becomes more gloomy. The fanciful pattern of figure 2 (where we compared with the LP-benchmark) now reappears. Running through the θ-scale, we switch from dominance to non-dominance.[6]

To get a better insight in the magnitude of disagreement between the different weighting methods, we finally present in Figure 5, for each value of θ, the percentage of S-Gini social welfare functions[7] where conflicts arise between the different weighting methods for the RP-benchmark.

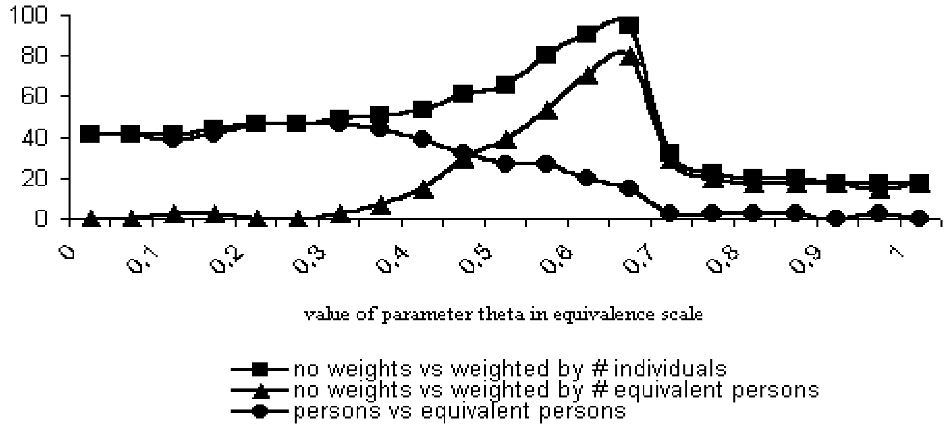

Figure 5: Percentage of δ between 1 and 5 for which a conflict arises (using RP benchmark)

[6] This result appears in a weaker form in the comparison with the LP-benchmark. With household weighting we find non-dominance throughout the whole θ-range. With individual weighting, we find dominance of the actual reform over the LP-benchmark throughout the whole θ-range. Only for weighting with equivalence scales, the pattern is mixed: everywhere non-dominance, interrupted by dominance of the actual reform at the values .55 and .60 for θ.

[7] With a distributional parameter δ between reasonable values of 1 and 5 in steps of 0.1. Changing the range and/or the grid does not alter the results.

The picture is revealing. The percentage of conflicts is largest between using households and one of the other methods, and a significant peak occurs for rather reasonable values of θ (around .6). Between the normatively more interesting weighting methods, the number of conflicts decreases with θ.

5. Conclusion

Given the impossibility of reconciling two normatively interesting properties, a choice between the corresponding weighting schemes is inevitable. The question we pose in our paper is whether the choice for one or another weighting scheme also matters empirically. We apply three different weighting methods to assess the 2000 reform proposal of personal income taxes in Belgium simulated via the microsimulation model SIRe of the Belgian Ministery of Finance.

We do find sensitivity of the results with respect to the different weighting methods. As a result, not weighting households is both normatively uninteresting and makes a difference empirically. In addition, using the number of equivalent individuals as weights to perform dominance analysis, leads to quite fanciful results with respect to the choice of equivalence scales. Hence, although this weighting method is normatively interesting, it might turn out to be ineffective. Needless to say that a profound statistical analysis is needed, to check whether our results will stand the test of statistical significance.

References

[1] Banks, J. and Johnson, P. (1994), Equivalence scale relativities revisited, *The Economic Journal* 104, 883-890.

[2] Bishop, J., Chow, K., Formby, J. and Ho, C. (1997), Did tax reform reduce actual US progressivity? Evidence form the taxpayer compliance measurement program, *International Tax and Public Finance* 4, 177-197.

[3] Burkhauser, R.V., Smeeding, T.M. and Merz, J. (1996), Relative inequality and poverty in Germany and the United States using alternative equivalence scales, *Review of Income and Wealth* 42(4), 381-400.

[4] Cowell, F.A. and Mercader-Prats, M. (1999), Equivalence of scales and inequality, DARP discussion paper 27, London School of Economics.

[5] Donaldson, D. and Weymark, J.A. (1983), Ethically flexible Gini indices for income distributions in the continuum, *Journal of Economic Theory* 29, 353-358.

[6] Ebert, U. (1995), Income inequality and differences in household size, *Mathematical Social Sciences* 30, 37-55.

[7] Ebert, U. (1997), Social welfare when needs differ: An axiomatic approach, *Economica* 64, 233-244.

[8] Ebert, U. (1999), Using equivalent income of equivalent adults to rank income distributions, *Social Choice and Welfare* 16, 233-258.

[9] Glewwe, P. (1991), Household equivalence scales and the measurement of inequality: Transfers from the poor to the rich could decrease inequality, *Journal of Public Economics* 44, 211-216.

[10] Howes, S. (1996), The influence of aggregation on the ordering of distributions, *Economica* 63, 253-272.

[11] Jenkins, S.P. and Cowell, F.A. (1994), Parametric equivalence scales and scale relativities, *Economic Journal* 104, 891-900.

[12] Lambert, P.J. (2001), *The Distribution and Redistribution of Income. A Mathematical Analysis*, 3rd edition, Manchester: Manchester University Press.

[13] Shorrocks, A.F. (1983), Ranking income distributions, *Economica* 50, 3-17.

[14] Shorrocks, A.F. (1995), Inequality and welfare evaluation of heterogeneous income distributions, Discussion Paper 447, University of Essex.