Società italiana di economia pubblica

WORKING PAPER
No 602
Rev1-June 2008

ON DETERMINING "CLOSE EQUALS GROUPS" IN DECOMPOSING REDISTRIBUTIVE AND RERANKING EFFECTS

ACHILLE VERNIZZI

Dipartimento di Scienze Economiche, Aziendali e Statistiche
Università degli Studi di Milano
E-mail: achille.vernizzi@unimi.it

SIMONE PELLEGRINO

Dipartimento di Scienze Economiche e Finanziarie "G. Prato"
Università degli Studi di Torino
E-mail: spellegrino@gmail.com

JEL Classification Numbers: H23; H24.

Keywords: Personal Income Tax, Redistributive Effect, Horizontal Inequity, Reranking.

ON DETERMINING “CLOSE EQUALS GROUPS” IN DECOMPOSING REDISTRIBUTIVE AND RERANKING EFFECTS

Achille Vernizzi

Dipartimento di Scienze Economiche, Aziendali e Statistiche

Università degli Studi di Milano

E-mail: achille.vernizzi@unimi.it

Simone Pellegrino

Dipartimento di Scienze Economiche e Finanziarie “G. Prato”

Università degli Studi di Torino

E-mail: spellegrino@gmail.com

27/11/2008

Abstract

Recently van De Van, Creedy and Lambert (2001) and Urban and Lambert (2008) have reconsidered the original Aronson, Johnson and Lambert (1994) decomposition of the redistributive effect in order to identify the optimal bandwidth that should be used in decomposing the redistributive effect when groups with close pre-tax incomes are considered. The methodology proposed by van De Van, Creedy and Lambert (2001) suggests choosing as the optimal bandwidth the one which maximizes the ratio between the potential effect V (which depends on the bandwidth) and the actual redistributive effect RE (which is invariant). Urban and Lambert (2008) discuss a set of further possible decompositions of the redistributive effect together with a decomposition of the Atkinson-Plotnick-Kakwani index into three terms. In this paper we want to throw some more light on the behavior of three of the main decompositions analyzed by Urban and Lambert (2008) in order to look for criteria to choose a bandwidth which allows the three different definitions of potential redistributive effect to be assumed as coherent as possible values and, in the meanwhile, to catch as much as possible of the potential vertical effect. We suggest looking for the bandwidth where the ratio between the maximum distance among the different potential vertical effect definitions and the minimum among the different potential vertical effects is minimum.

JEL Classification Numbers: H23; H24.
Keywords: Personal Income Tax, Redistributive Effect, Horizontal Inequity, Reranking.

1. Introduction

Decomposing redistributive effect across groups of pre-tax equals into vertical, horizontal and reranking effect has been intensively studied in recent years. The original work by Aronson, Johnson and Lambert (1994), hereafter AJL, considers exact pre-tax equals in portioning the pre-tax income distribution.

As van de Ven, Creedy and Lambert (2001), hereafter VCL, pointed out, in taxation this is not the casein reality: only groups with close pre-tax incomes can be considered. They overcame this problem in order to identify the optimal bandwidth that should be used in decomposing the redistributive effect. Therefore, VCL methodology suggests choosing as the optimal bandwidth the one which maximizes the ratio between the potential effect (which depends on the bandwidth) and the actual redistributive effect (which is invariant). Here a problem arises as this ratio may have more than one relative maximum and presents a layout which may be irregular, and somewhere quite irregular; as a consequence, identifying univocally the maximum is not so obvious (Vernizzi and Pellegrino 2007).

Urban and Lambert (2008), hereafter UL, present an exhaustive discussion on a complete set of possible redistributive effect decompositions, and introduce new indexes based on the taxation of close equals by their average tax rate.

In this work we would like to contribute to VCL's and UL's discussions with some suggestions about the choice of a convenient bandwidth by intensively looking at the empirical analysis. We conclude by looking for the bandwidth (or for a set of bandwidths) where the ratio between the maximum distance among the different potential vertical effect definitions and the minimum among the different potential vertical effects is minimum.

The structure of the paper is as follows. In Section 2 we recall how the original AJL decomposition should be applied in the real world where strict equals groups are rare and, consequently, they must be replaced by "close equals" groups; to overcome this problem, according to UL's suggestions, alternative *RE* decompositions are introduced together with the decomposition of the Atkinson-Plotnick-Kakwani index (hereafter R^{APK})[1]. In Section 3 we report the values the indexes assume at bandwidth limits, that is either when the bandwidth tends to zero or when it covers the whole income distribution range; then we outline some preliminary a priori considerations about some aspects of their behavior. The empirical behavior of indexes is analyzed in Section 4. Section 5 discusses whenever a bandwidth with "optimal" or at least "desirable" properties can be identified. Section 6 concludes.

2. Redistribution and reranking indexes

Let G_y and G_{y-T} be the Gini index on the gross and net incomes respectively. The redistributive index *RE* is equal to $RE = G_y - G_{y-T}$. It is well-known that the Gini coefficient fails to decompose across subgroups into, between and within group inequality components in case subgroup income ranges overlap. When considering the pre-tax income parade, if groups are selected in a sequential order, so that a pooling of all groups incomes is in a non decreasing order, we have that $G_y = G_y^B + G_y^W$, where G_y^B

[1] The decomposition of R^{APK} is described and discussed in Urban and Lambert (2005); for further details see Vernizzi (2007).

is the between-group Gini pre-tax index and G_y^W is the within-group Gini pre-tax index[2]. However, if post-tax income groups contain the same subjects they did before taxation, it is no longer granted that the after tax maximum value in the *i*-th group is not greater than the minimum value in the $i+1$-th group and that no intersection (or overlapping) effect appears among groups.

If taxation induces overlapping among groups, the post-tax Gini index becomes $G_{y-T} = G_{y-T}^B + G_{y-T}^W + G_{y-T}^t$, where [3] $G_{y-T}^t = R^{AJL} = G_{y-T} - \left(G_{y-T}^B + G_{y-T}^W \right)$.

When exact equals are considered

In their seminal paper, AJL not only organize groups so that no overlapping effect exists for pre-tax groups, but also implicitly assume that for the after-tax income parade (*i*) the group averages maintain the same ranking as before taxation and (*ii*) the within group orderings remain the same as before taxation. If this is the case, the post-tax concentration index (evaluated when post-tax incomes are ordered according to the

[2] G_y^B is the Gini index for pre-tax incomes when within each group all incomes are substituted by their group average; $G_y^W = \sum_k a_{k,y} G_{k,y}$, where $G_{k,y}$ is the Gini index for the *k*-th group and $a_{k,y}$ is the product of the *k*-th group population share and pre-tax income share.

[3] G_{y-T}^B and $G_{y-T}^W = \sum_k a_{k,y-T} G_{k,y-T}$ are the analog forms for G_y^B and G_y^W when incomes have been taxed; in particular $a_{k,y-T}$ is the product of the *k*-th group population share and post-tax income share. G_{y-T}^t is what Dagum (1997) calls "the transvariation term". In UL notation $R^{AJL} = G_{y-T} - D_4$, where $D_4 = G_{y-T}^B + G_{y-T}^W$ is the concentration index for the after tax income parade, ordered according to non decreasing group averages and, within each group, in a non decreasing order. The relations which involves Gini and concentration indexes components are analyzed, e.g., in Vernizzi (2007).

order they had before taxation) is $\left(G^{B}_{y-T}+G^{W}_{y-T}\right)$, so that $R^{AJL}=R^{APK}$, being R^{APK} the Atkinson-Plotnick-Kakwani reranking index[4].

If we split G_y and G_{y-T} into the above described components, as AJL do, the redistributive effect can be written as $RE=\left(G^{B}_{y}-G^{B}_{y-T}\right)-\left(G^{W}_{y-T}-G^{W}_{y}\right)-R^{APK}$. A further simplification can be applied when the analysis is limited to the case in which the population groups contain exact pre-tax equals, which implies $G^{W}_{y}=0$ and $G^{B}_{y}=G_y$. In this case the redistributive effect can be expressed as $RE=\left(G_y-G^{B}_{y-T}\right)-G^{W}_{y-T}-R^{APK}$.

AJL name $\left(G_y-G^{B}_{y-T}\right)$ the vertical potential redistributive effect: it loses part of its potentiality whenever either the within-group inequality index G^{W}_{y-T} or the group overlapping index $R^{AJL}=G^{t}_{y-T}=G_{y-T}-\left(G^{B}_{y-T}+G^{W}_{y-T}\right)=R^{APK}$ becomes different from zero after taxation.

When close equals are considered

However, as observed before, this decomposition can be correctly applied provided that each group is composed by subjects with the same pre-tax income and taxation does not modify either the ranking among group averages or the within-group rankings (van de Ven, Creedy and Lambert, 2001; Vernizzi, 2006; Urban and Lambert, 2008).

[4] In UL notation D_1 is the concentration index for after-tax incomes, when ordered according to the before taxation ranking. D_1 may be different from $D_4=G^{B}_{y-T}+G^{W}_{y-T}$ and, in general, it is. In our notation D_1 is D_{y-T}.

In the real world, even for gross incomes the within-group Gini index, G_y^W is generally different from zero, as only groups with close pre-tax incomes can be considered. As a consequence, only bandwidths of income containing close-equals must be chosen.

More in general, neither post-tax grouping means maintaining the same order they had for the pre-tax income parade nor, that, within each group, the order of the incomes remains unchanged in the transition from the pre- to the post-tax incomes; in this case the residual of the *RE* decomposition is generally not equal to the APK index, which can be more generally defined as $R^{APK} = G_{y-T} - D_{y-T} = G_{y-T} - \left(D_{y-T}^B + D_{y-T}^W\right)$. D_{y-T} is the concentration index for the post-tax income parade when incomes are ranked according the pre-tax income non-decreasing ranking; D_{y-T}^B and D_{y-T}^W are, respectively, the between and the within group concentration indexes for post-tax income parade[5]. We can confirm these violations using a SHIW dataset, even if the magnitude of these "unpleasant" outcomes depends on the income range (bandwidth) chosen for each group. It is worth stressing that, according to empirical evidence[6], the income bandwidth acts in opposite directions towards group reranking and within-group reranking: the larger the bandwidth is, the less probable is the former and the more frequent the latter.

---

[5] D_{y-T}^B is defined as the concentration index when all incomes inside each group are substituted by the group income average and, moreover, groups are ordered according to pre-tax group averages. $D_{y-T}^W = \sum_k a_{k,y-T} D_{k,y-T}$, where $D_{k,y-T}$ is the concentration index for the *k*-th group, when after tax incomes are ordered according to their pre-tax order, and $a_{k,y-T}$ is the product of the *k*-th group population share and post-tax income share.

[6] Lambert and Urban (2005), Vernizzi and Pellegrino (2007).

In addition, as the bandwidth increases, G_y^W can be no more close to zero, so that the redistributive effect can be no more evaluated as $RE = \left(G_y - G_{y-T}^B\right) - G_{y-T}^W - R^{APK} = \left(G_y^B - G_{y-T}^B\right) - G_{y-T}^W - R^{APK}$; it becomes more realistic to turn back to the more complete decomposition

$$RE = \left(G_y^B - G_{y-T}^B\right) - \left(G_{y-T}^W - G_y^W\right) - R^{AJL} = V^{VCL} - H^{VCL} - R^{AJL} \tag{1}$$

having defined $V^{VCL} = \left(G_y^B - G_{y-T}^B\right)$ and $H^{VCL} = G_{y-T}^W - G_y^W$.

When using the above decomposition, one returns to the idea of constituting close-equals groups, and focuses on the eventual enlargement of the within-group inequality $\left(G_{y-T}^W - G_y^W\right) = H^{VCL}$ term, together with the group overlapping term R^{AJL}, to measure the loss in potential vertical redistribution effect which is measured by $\left(G_y^B - G_{y-T}^B\right) = V^{VCL}$.

UL present other *RE* decompositions which hold also when groups do not include just equals or between or within groups rerankings introduced by taxation. Here we shall consider two of these decompositions, both of them apply the idea of smoothed taxation within group, which is introduced by UL in coherence with the principle of close equals groups: if groups contain close equals, their incomes should be taxed by a same tax rate, which can be properly estimated by the group average tax rate. After having applied a same tax rate to all incomes in group *k*, the Gini index for group *k* remains exactly equal to the pre-tax $G_{k,y}$; however the smoothed within group Gini index $G_{y-T}^{SW} = \sum_k a_{k,y-T} G_{k,y}$ is generally different from $G_y^W = \sum_k a_{k,y} G_{k,y}$, because in general $a_{k,y} \neq a_{k,y-T}$.

UL define $RE = V^{AJL} - H^{AJL} - R^{AJL}$, where $V^{AJL} = G_y - \left(G^{B}_{y-T} + G^{SW}_{y-T}\right)$ and $H^{AJL} = G^{W}_{y-T} - G^{SW}_{y-T}$, so that:

$$RE = V^{AJL} - H^{AJL} - R^{AJL} = \left(G_y - G^{B}_{y-T} - G^{SW}_{y-T}\right) - \left(G^{W}_{y-T} - G^{SW}_{y-T}\right) - R^{AJL} \qquad (2)$$

In expression (2) the potential vertical effect is measured by the difference between the pre-tax Gini index and the Gini index for an artificial post tax income parade, which, by constructions, excludes any group overlapping[7].

The "pure" horizontal inequity is measured by the enlargement of within group inequality, with respect to what would induce a smoothed taxation; group overlapping introduced by taxation, is measured by R^{AJL} as in equation (1).

Both expressions (1) and (2) take into account only a part of horizontal inequity, eventually introduced by a taxation system, in fact the two *RE* decompositions do not consider within group and between group eventual rerankings. Actually, the Atkinson-Plotnick-Kakwani index R^{APK} can be decomposed into three terms[8]: $R^{APK} = R^{AJL} + R^{EG} + R^{WG}$. Together with the overlapping term R^{AJL} which has been already described, there are two further terms: the former, R^{EG}, measures group

[7] UL define V^{AJL} and H^{AJL} in an apparently different way. They define D_5 and D_6 as concentration indexes calculated on smoothed net incomes: the D_5 index ranks groups according to the same order they had before taxation, even if the taxation changed the income average order among groups; the D_6 index ranks groups that are ranked in a non decreasing order with respect to their post-tax average incomes. D_3 is the concentration index for (non-smoothed) after tax incomes, when groups follow the same order as before taxation, whilst within group incomes are in non decreasing order; then $V^{AJL} = G_y - D_6$ and $H^{AJL} = D_4 - D_6 = D_3 - D_5$.

[8] Lambert and Urban (2005). See also Vernizzi (2006) for analytical details.

averages reranking, whilst R^{WG} measures the reranking effect due to within groups reshuffling. More in detail [9] $R^{EG} = G^{B}_{y-T} - D^{B}_{y-T}$ and $R^{WG} = G^{W}_{y-T} - D^{W}_{y-T}$.

The latter UL decomposition we consider is $RE = V - H - R^{APK}$, where $V = G_y - \left(D^{B}_{y-T} + G^{SW}_{y-T}\right)$ and $H = D^{W}_{y-T} - G^{SW}_{y-T}$. Then:

$$RE = V - H - R^{APK} = \left(G_y - D^{B}_{y-T} - G^{SW}_{y-T}\right) - \left(D^{W}_{y-T} - G^{SW}_{y-T}\right) - R^{APK} \qquad (3)$$

UL notice that decomposition (3) has the advantage of synthesizing the whole information set into one equation[10]. Table 1 summarizes Gini and concentration indexes definitions.

TABLE 1 ABOUT HERE

What decomposition is more suitable to analyze the redistributive effect and what bandwidth should be chosen is a problem not definitely solved: VCL suggest choosing a bandwidth where $\left(V^{VCL}/RE\right)$ is maximum. This ratio may have more than one relative maximum and presents a layout which may be irregular, and somewhere quite irregular; as a consequence, identifying univocally the maximum is not so obvious. We got over this problem.

---

[9] UL define $R^{EG} = D_4 - D_3$ and $R^{WG} = D_3 - D_1$.

[10] UL define *V* and *H*, respectively, as $V = G_y - D_5$ and $H = D_1 - D_5$.

3. A priori considerations on indexes behavior

On a priori considerations, we can easily state the values that the indexes considered here assume at bandwidth limits, that is either when bandwidth tends to zero or the maximum available range (Table 2).

TABLE 2 ABOUT HERE

When the bandwidth tends to zero[11], $V^{VCL} = V^{AJL} = RE$, $V = G_y - D_{y-T}$ (the Reynolds-Smolenky total redistribution index) and $H^{VCL} = H^{AJL} = H = 0$; it follows that at bandwidth zero $V \geq V^{VCL} = V^{AJL}$. Conversely, when the bandwidth is maximum, that is equal to the observed income range, $V^{VCL} = V^{AJL} = V = 0$, $H^{VCL} = H^{AJL} = -RE$ and $H = -\left(G_y - D_{y-T}\right)$, so that when the bandwidth coincides with the maximum range $H \leq H^{VCL} = H^{AJL}$. In relation to the reranking effects, we have that when the bandwidth is zero $R^{AJL} = R^{WG} = 0$ and $R^{EG} = R^{APK}$, whilst for maximum bandwidth $R^{AJL} = R^{EG} = 0$ and $R^{WG} = R^{APK}$.

The difference $V - V^{AJL}$ is equal to R^{EG}, which being non-negative, implies V to dominate V^{AJL}. Less evident is the relation between V^{VCL} and V^{AJL} and, especially, that between V and V^{VCL}: in fact $V^{VCL} - V^{AJL} = G_{y-T}^{SW} - G_y^{W}$ and $V - V^{VCL} = R^{EG} - \left(G_{y-T}^{SW} - G_y^{W}\right)$.

[11] A bandwidth equal or tending to zero implies that each subject is considered by itself.

In order to throw some light on these relations, we recall how G_y^W and G_{y-T}^{SW} can be represented as weighted sums of average absolute differences, calculated within each group:

$$
\begin{aligned}
G_y^W &= \frac{1}{n^2\mu}\sum_{i=1}^{K} G_{i,y} n_i^2 \mu_i = \frac{1}{2n^2\mu}\sum_{i}^{K} \Delta_{i,y} n_i^2 \\
G_{y-T}^{SW} &= \frac{1}{n^2\mu\left(1-\overline{t}\right)}\sum_{i=1}^{K} G_{i,y} n_i^2 \mu_i \left(1-t_i\right) = \frac{1}{2n^2\mu\left(1-\overline{t}\right)}\sum_{i}^{K} \Delta_{i,y}\left(1-t_i\right) n_i^2
\end{aligned}
\tag{4}
$$

where μ is the average income for all the subjects considered in the sample, n is the number of equivalent subjects in the sample, n_i is the number of equivalent subjects [12] in group k, t_i is the i-th group average tax rate, $\overline{t}$ is average tax rate for the whole sample, $\Delta_{i,y} = \left(2/n_i^2\right)\sum_{s=1}^{k_i}\sum_{h>s}^{k_i}\left(y_{i,h} - y_{i,s}\right) n_{i,s} n_{i,h}$, k_i being the number of cases registered in group i and $n_{i,s}$ the weight associated to income y_i; within each group incomes are ranked in a non decreasing order. We can then write:

$$
\begin{aligned}
G_{y-T}^{SW} - G_y^W &= \frac{1}{n^2\mu\left(1-\overline{t}\right)}\sum_{i=1}^{K} G_{i,y} n_i^2 \mu_i \left(\overline{t} - t_i\right) = \\
&= \frac{1}{2n^2\mu\left(1-\overline{t}\right)}\sum_{i=1}^{K} \Delta_{i,y}\left(\overline{t} - t_i\right) n_i^2
\end{aligned}
\tag{5}
$$

Due to the asymmetry of income distributions, which makes bandwidths in the left tail more crowded than those in the right tail, in (5) positive $\left(\overline{t} - t_i\right)$'s, are likely to receive a weight more than proportional than the negative $\left(\overline{t} - t_i\right)$'s; if this is the case, V^{VCL} is expected to be not lower than V^{AJL} [13].

---

[12] A sum of equivalent subjects may be a non integer number.

[13] For instance when the income range is split into two groups, $\mu_2 > \mu_1$, each having the same spread but not necessarily the same number of subjects, as $\overline{t} = \left(t_1\mu_1 n_1 + t_2\mu_2 n_2\right)/\left(\mu_1 n_1 + \mu_2 n_2\right)$, $t_2 > t_1$, we can write

Turning to V and V^{VCL}, the sign of the difference $V - V^{VCL}$ depends on the difference $R^{EG} - \left(G_{y-T}^{SW} - G_{y}^{W}\right)$, where $G_{y-T}^{SW} - G_{y}^{W}$ is likely to be non negative, due to the above considerations, and R^{EG} is certainly non-negative. So, on a priori considerations, we can only conclude that for the bandwidth tending to zero $V - V^{VCL}$ has R^{APK} as its limit, and for the bandwidth tending to the maximum range, $V - V^{VCL}$ has zero as its limit.

In relation to the horizontal loss measures, we observe that $H^{AJL} - H$ is equal to R^{WG}, the within group reranking index[14], which is non-negative and that the difference between $H^{VCL} - H$ is equal to the sum of R^{WG} and $\left(G_{y-T}^{SW} - G_{y}^{W}\right)$: as the former is always non-negative and the latter is likely to be non negative, we expect that $H^{AJL} \geq H$; moreover given that $\left(G_{y-T}^{SW} - G_{y}^{W}\right)$ is non-negative, H^{VCL} should be not lower than H^{AJL}, so that, summarizing, we expect that $H^{VCL} \geq H^{AJL} \geq H$, where the second inequality always applies.

4. Empirical analysis

In this section we investigate by an empirical analysis how the group bandwidth influences the components of the redistribution index V^{VCL}, H^{VCL} and R^{AJL} in equation

---

$G_{y-T}^{SW} - G_{y}^{W} = \left\{1\big/\left[n^2\mu\left(1-\bar{t}\right)\right]\right\}\cdot\left[G_{1,y}n_1^2\mu_1\left(\bar{t}-t_1\right) + G_{2,y}n_2^2\mu_2\left(\bar{t}-t_2\right)\right] =$
$= \left\{\left(n_1\mu_1 n_2\mu_2\right)\big/\left[n^2\mu\left(1-\bar{t}\right)\left(n_1\mu_1 + n_2\mu_2\right)\right]\right\}\cdot\left[G_{1,y}n_1\left(t_2-t_1\right) - G_{2,y}n_2\left(t_2-t_1\right)\right]$, which is greater than zero if $\left(n_1/n_2\right) > \left(G_{2,y}/G_{1,y}\right)$. This is quite likely to be verified for an income asymmetric distribution.

14 $R^{WG} = G_{y-T}^{W} - D_{y-T}^{W} = \frac{1}{2n^2\mu\left(1-\bar{t}\right)}\sum_{i=1}^{K}\left(G_{i,y-T} - D_{i,y-T}\right)n_i^2\mu_i\left(1-t_i\right) = \frac{1}{n^2\mu\left(1-\bar{t}\right)}\sum_{i=1}^{K}R_i^{APK}n_i^2\mu_i\left(1-t_i\right)$, having defined with R_i^{APK} the Atkinson-Plotnick-Kakwani for the i-th group.

(1), V^{AJL} and H^{AJL} in equation (2), and V in equation (3), together with the components of the Atkinson-Plotnick-Kakwani index $R^{APK} = R^{AJL} + R^{EG} + R^{WG}$.

As stated before, our aim is to contribute to the discussion about the choice of a proper bandwidth: a proper bandwidth should catch as much as possible of the potential redistributive effect and, in the meanwhile, should get as close as possible measures from the three indexes V^{VCL}, V^{AJL} and V.

Our experiment was conducted on the basis of the Bank of Italy survey on household incomes and wealth (SHIW). The 2004 Italian SHIW dataset provides demographic and post-tax income microdata for a representative cross-section of 12,713 taxpayers and 8,012 households (20,581 individuals). These data were used to obtain gross and net incomes according to the Italian Personal Income Tax (Pellegrino, 2007b). In order to deal in some way with two different data bases, the experiment was conducted with respect to both individual and family equivalent incomes. Equivalent incomes were obtained by dividing total family incomes by an equivalence scale; the scale here adopted is the Cutler scale which can be expressed as $CS_h = \left(NA_h + \alpha NC_h\right)^{\beta}$, having (arbitrarily) set $\alpha = 0.5$ and $\beta = 0.65$. Ebert and Moyes (2000) observe that, in applying equivalence scales, the choice of the weight may be arbitrary: we consequently decided to weigh equivalent incomes by the lower and the upper boundaries, the former being 1[15] and the latter being the component number associated to each family[16]. Once the 2004 gross income parade was obtained, the 2006 and 2007 distributions were estimated considering the impact of the inflation rate (Pellegrino, 2007c). We found that

[15] $\alpha = \beta = 0$.

[16] $\alpha = \beta = 1$.

the results are quite analogous for both weights equal to family components and for all weights equal to 1; moreover, the results are also very similar across different years, so that, for the sake of simplification, here we only report results referred to 2004, for individuals and for household equivalent incomes - weight 1[17]. Figures 1a and 1b show the behavior of the three potential redistributive effects, V^{VCL}, V^{AJL} and *V*, are plotted together with the constant line of the Reynolds-Smolenky total redistributive effect.

The three indexes which measure the loss in horizontal equity, H^{VCL}, H^{AJL} and *H*, are reported in Figures 2a and 2b, together with R^{AJL} and R^{APK}, the latter being constant; all the above measures are expressed as percentages of the redistributive effect *RE*.

The decomposition of R^{APK} is represented in Figures 3a and 3b: R^{AJL}, R^{EG} and R^{WG} are there expressed as percentages of R^{APK}.

As we noticed in the previous section, in correspondence of a zero bandwidth, both V^{VCL} and V^{AJL} are equal to *RE*, whilst *V* is equal to $G_y - D_{y-T}$, the Reynolds-Smolensky redistribution index, which is greater than *RE*. For our minimum bandwidth, which is 10 euro, in Figure 1 V^{VCL} and V^{AJL} show a 0.7 % increase (0.8 % when dealing with family equivalent incomes: Figure 1b) with respect to the limit value for the bandwidth tending to zero, that is *RE*. In Figure 1 the two lines are not distinguishable and show a steep ascent up to bandwidths of around 300 euro; then V^{AJL} leaves V^{VCL} and becomes indistinguishable from *V* for bandwidths larger than 400, when considering individuals, and larger than 550-600 when considering family equivalent incomes. *V* shows a decreasing trend; before becoming indistinguishable

[17] Even if limited to V^{VCL}, H^{VCL}, R^{AJL}, R^{EG} and R^{WG}, Vernizzi and Pellegrino (2007) report all graphs for the three tax systems (2004, 2004 and 2006), concerning both individuals and family equivalent incomes (weight 1 and weight equal to family components).

from V^{AJL} it dominates V^{VCL} and V^{AJL}, then the V line crosses V^{VCL} and continues descending together with V^{AJL}, leaving the V^{VCL} line above. When the bandwidth is 3,000 euro, for individuals the three lines are still greater than *RE*: V^{VCL} is almost 1.0% greater than *RE* (1.2% families), V^{AJL} and V are only 0.4% greater than *RE* (0.6% families).

Even if our analysis tries to focus on smaller bandwidths than UL do, our findings are substantially consistent with UL results; what appears to be different is that the lines presented by UL look much more regular than the ones represented here. Our lines are the more irregular the more they depart from the axes origin: the irregularities are more similar to irregular waves than to completely random white noises.

More in detail, we observe that:

(*i*) $V^{VCL} \simeq V^{AJL}$ as long as $G_{y-T}^{SW} \simeq G_y^W$; V^{VCL} becomes greater than V^{AJL} when G_{y-T}^{SW} becomes sensibly greater than G_y^W;

(*ii*) $V^{AJL} \simeq V$ after R^{EG} becomes $\simeq 0$; as long as R^{EG} is not negligible $V^{AJL} < V$;

(*iii*) $V^{VCL} \simeq V$ for bandwidths where $R^{EG} \simeq \left(G_{y-T}^{SW} - G_y^W\right)$; $V^{VCL} < V$ as long as $R^{EG} > \left(G_{y-T}^{SW} - G_y^W\right)$ and, conversely, $V^{VCL} > V$ after R^{EG} becomes lower than $\left(G_{y-T}^{SW} - G_y^W\right)$.

FIGURES 1a-1a bis, 1b, 1b bis, 2a, 2b, ABOUT HERE

Figures 2a and 2b represent the behavior of the three indexes which measure the horizontal effect, together with R^{AJL}, the overlapping index, and R^{APK}, the global reranking index; as expected, the three indexes here considered, H^{VCL}, H^{AJL} and H, assume a value which is very close to zero when the bandwidth is 10 euro. H presents few and insignificant positive values just for the tiniest bandwidth; then it starts a descending trend towards the limit value $-\left(G_y - D_{y-T}\right)$. Conversely in correspondence of bandwidths of 10-3,000, considered here, H^{VCL} and H^{AJL} always present positive values. In particular when the bandwidth is 3,000 euro, H^{VCL} looks to be still increasing, whilst H^{AJL} has already started the descending trend. Similarly to V^{VCL}, V^{AJL} and V, when bandwidths become large, H^{VCL}, H^{AJL} and H present relatively strong irregularities.

FIGURES 3a-3b ABOUT HERE

The decomposition of R^{APK} is represented in Figures 3a and 3b: R^{AJL}, R^{EG} and R^{WG} are expressed as percentages of R^{APK}. R^{AJL}, which is zero both at bandwidth zero and at bandwidth maximum, shows a quite asymmetric line (as could be noted also from Figures 2a-2b, where it has just been rescaled by R^{APK}/RE): for individuals, at 10 euro bandwidth it has already jumped up to 67% of R^{APK} (58% for families) and it reaches its maximum value, 88%, at the 100 euro bandwidth (86% for families) then it begins to descend and at a 3,000 euro bandwidth it is roughly at 25% of R^{APK}. R^{EG}, which coincides with it is R^{APK} when the bandwidth is a point bandwidth, is 32,4% of R^{APK} at the 10 euro bandwidth (40% for families); it decreases quite soon and at a 300 euro bandwidth is already less than 1% of R^{APK}. R^{WG} appears to be a direct function of the bandwidth, even if at decreasing rates: at a 3,000 bandwidth it is nearly 80% of R^{APK}

(70% for families). Similarly to what happens for the potential vertical indexes and the horizontal iniquity indexes, as bandwidths become larger, R^{AJL} and R^{WG} present relatively strong irregularities: this does not happen for R^{EG}, due to the fact that this index is quite low for large bandwidths.

Now let us investigate H^{VCL}, H^{AJL} and H behavior; we can represent H^{VCL} as:

$$
\begin{aligned}
H^{VCL} = G^{W}_{y-T} - G^{W}_{y} &= \sum_{i=1}^{K}\left(G_{i,y-T}\frac{n_i^2\mu_i\left(1-t_i\right)}{n^2\mu\left(1-\bar{t}\right)} - G_{i,y}\frac{n_i^2\mu_i}{n^2\mu} \right) = \\
&= \sum_{i=1}^{K}\left\{ \frac{n_i^2\mu_i}{n^2\mu\left(1-\bar{t}\right)}\left[G_{i,y-T}\left(1-t_i\right) - G_{i,y}\left(1-\bar{t}\right)\right]\right\} \qquad (6)\\
&= \frac{1}{2n^2\mu\left(1-\bar{t}\right)}\sum_{i=1}^{K}\left[\Delta_{i,y-T} - \Delta_{i,y}\left(1-\bar{t}\right)\right] n_i^2
\end{aligned}
$$

and H^{AJL} as:

$$
\begin{aligned}
H^{AJL} = G^{W}_{y-T} - G^{SW}_{y-T} &= \sum_{i=1}^{K}\left(G_{i,y-T} - G_{i,y}\right)\frac{n_i^2\mu_i\left(1-t_i\right)}{n^2\mu\left(1-\bar{t}\right)} = \\
&= \frac{1}{2n^2\mu\left(1-\bar{t}\right)}\sum_{i=1}^{K}\left[\Delta_{i,y-T} - \Delta_{i,y}\left(1-t_i\right)\right] n_i^2
\end{aligned} \qquad (7)
$$

For lower income groups, where the tax rate for each subject may be much lower than $\bar{t}$, $\Delta_{i,y-T}$ may be greater than $\Delta_{i,y}\left(1-\bar{t}\right)$, even if $\Delta_{i,y-T} \le \Delta_{i,y}$ which may cause H^{VCL} to result positive: in fact due to the asymmetry of income distributions, it is likely that lower income intervals contain more subjects than higher income groups do, so that the weighed sum represented in (6) may result to be positive even if $\Delta_{i,y-T} \le \Delta_{i,y}, \forall k$. Then we can conclude that the H^{VCL} remains positive until bandwidths are large enough to make a sufficient number of $\Delta_{i,y-T}$'s – especially in the left hand side of the distribution - small enough to be less than their corresponding pre-tax $\Delta_{i,y}$, multiplied by $\left(1-\overline{t}\right)$[18].

When dealing with incomes in the right distribution queue, the contrary happens, but, due to the distribution asymmetry, in the left hand tail income groups generally present weights greater than those in the right hand tail.

Turning now to H^{AJL}, as for lower incomes $\left(1-t_i\right) \geq \left(1-\overline{t}\right)$, in the left distribution queue, the relation $\Delta_{i,y-T} \leq \Delta_{i,y}\left(1-t_i\right)$ is more likely to be verified than the relation $\Delta_{i,y-T} \leq \Delta_{i,y}\left(1-\overline{t}\right)$. This consideration should explain why H^{AJL} starts decreasing earlier than H^{VCL}. In any case it is excluded that H^{AJL} is positive when all groups' post-tax Gini indexes are lower than the corresponding pre-tax ones.

Let us now define $\Delta^{D}_{i,y-T} = \left(2/n_i^2\right)\sum_{s=1}^{k_i}\sum_{h>s}^{k_i}\left(\tilde{y}_{i,h} - \tilde{y}_{i,s}\right)n_{i,s}n_{i,h}$, where $\tilde{y}_{i,h}$ is net income for subject *h* in group *i*, as incomes are here ordered according to pre-tax ranking; we can now express *H* as:

$$
\begin{aligned}
H = D^{W}_{y-T} - G^{SW}_{y-T} &= \sum_{i=1}^{K}\left(D_{i,y-T} - G_{i,y}\right)\frac{n_i^2\mu_i\left(1-t_i\right)}{n^2\mu\left(1-\overline{t}\right)} = \\
&= \frac{1}{2n^2\mu\left(1-\overline{t}\right)}\sum_{i=1}^{K}\left[\Delta^{D}_{i,y-T} - \Delta_{i,y}\left(1-t_i\right)\right]n_i^2
\end{aligned}
\qquad (8)
$$

Looking at Figures 2a and 2b we realize that $\Delta_{i,y}\left(1-t_i\right)$ becomes greater than $\Delta^{D}_{i,y-T}$ even when bandwidths are not so large[19], at least for the left distribution queue where

[18] We observe also that H^{VCL} may be positive even when all groups post-tax Gini indexes are lower than the corresponding pre-tax ones, due two the different weight system: for lower incomes after tax weights should in fact be higher then the corresponding pre-tax ones and the reverse should hold for higher incomes.

groups are more crowded, and, consequently, receive a weight which is heavier than in the right one.

Table 3 reports the values for RE and R^{APK} decompositions, evaluated at bandwidths 100, 200, 300, 400, 500, 600, 700 and 2,000; together with their standard errors obtained by 2,000 bootstrap replications. From the figures reported in the tables it results that the ratios between the indexes and their standard errors are generally quite high, except for those which concern $\left(R^{EG}/R\right)$. The ratios $\left(R^{EG}/R\right)/SE\left\{R^{EG}/R\right\}$ range from 6.78 to 8.13 when the bandwidth is 100 euro, from 0.75 to 2.20 when the bandwidth is enlarged up to 700 euro, and they are not greater than 1.36 when the bandwidth is 2000. It is worth stressing that the 95% bootstrap percentiles are generally quite similar to those calculated assuming normality except for those related to $\left(R^{EG}/R\right)$; this result is in line with UL findings. Their simulations lead to the conclusion that the distribution for R^{EG} is asymmetric while the distributions for the other indexes they consider are symmetric and, moreover, that the bootstrap estimated standard error for R^{EG} is almost twice than that of the true distribution. Then we can conclude that the point estimates for RE components should be quite reliable. The same should hold for R^{WG}/R and R^{AJL}/R; $\left(R^{EG}/R\right)$ remain apart, perhaps due its relatively small magnitude, but not only for this reason: R^{WG}/R is small when the bandwidth is 100 euro, nevertheless it shows lower standard errors and bootstrap confidence intervals

[19] We observe that $\Delta^{D}_{i,y-T} \leq \Delta_{i,y-T}$ (and, obviously, $D_{i,y-T} \leq G_{i,y-T}$) which helps to explain why H becomes negative much before than H^{AJL}.

more similar to those obtained by the normal distribution. Some caution should be adopted also for H when it assumes small absolute values.

Tables 3 and 4 ABOUT HERE

5. On determining an "optimal" bandwidth

According to VCL decomposition, V^{VCL} should eliminate both measurement errors and anomalous values by averaging within group incomes. VCL suggest choosing the bandwidth which maximizes the potential redistributive effect. On one side, it is true that the larger the groups, the more efficacious the smoothing action is; on the other side, the larger the groups, the less equals incomes are within groups. AJL methodology appears to be quite appealing for the horizontal effect measure adopted: as we stressed in the previous chapter, H^{AJL} cannot result in being positive when all groups' post-tax Gini indexes are lower than the corresponding pre-tax ones. It cannot be excluded at all for H^{VCL}. H presents the undoubted advantage of being considered together with R^{APK} and so not only with R^{AJL}. However its interpretation is not simple because negative values of H imply the horizontal effect is added and not subtracted to the overall redistributive effect.

Looking either at Figures 1a and 1b in the present paper or the corresponding UL figures, we can see that, in correspondence of some bandwidths, the three potential redistributive effects are quite close, or even coincide. We would observe even closer values when the bandwidth tends to cover the whole income range: in this case, however, as already noticed, the three indexes would tend to zero and would not capture any potential vertical redistribution at all.

Figures 4a and 4b represent R^{EG} and $\left(G_{y-T}^{SW} - G_y^W\right)$ (as percentages of *RE*) plotted on the bandwidth; as observed in Chapter 3, $R^{EG} = V - V^{AJL}$, $\left(G_{y-T}^{SW} - G_y^W\right) = V^{VCL} - V^{AJL}$ and $\left(V - V^{VCL}\right) = R^{EG} - \left(G_{y-T}^{SW} - G_y^W\right)$. The behavior shown by R^{EG} in our empirical analysis is confirmed also by UL analysis: when bandwidths become large, group average re-rankings annihilate, sooner or later, depending on tax fairness: for individuals R^{EG} becomes zero sooner than for families. If we go back to expression (5) we can see that $\left(G_{y-T}^{SW} - G_y^W\right)$ is totally negligible for small bandwidths, where $\left(n_i^2 / n^2\right)$ results to be quite small and their sum is much less than 1 when bandwidths are tiny and therefore little crowded. As bandwidths increase, a more than proportional increase in $\left(n_i^2 / n^2\right)$ is not compensated by a convergence of the t_i towards their average $\overline{t}$ and the difference between G_{y-T}^{SW} and G_y^W increases more than proportionally. Certainly, $\left(G_{y-T}^{SW} - G_y^W\right)$ is expected to decrease and to tend to zero as bandwidths enlarge, as $\lim_{b \to MAX} G_{y-T}^{SW} = \lim_{b \to MAX} G_y^W = G_y$. Moreover, as already observed, when their difference reaches zero, *V*, V^{VCL} and V^{AJL} too become very small and tend to zero.

In the bandwidth range considered here, it follows that where $R^{EG} = \left(G_{y-T}^{SW} - G_y^W\right)$, (*i*) *V* is equal to V^{VCL} and (*ii*), the greater of the two distances $V - V^{AJL} = R^{EG}$, $V^{VCL} - V^{AJL} = \left(G_{y-T}^{SW} - G_y^W\right)$ is at its minimum value.

For individuals, at bandwidth 280 where R^{EG} and $G_{y-T}^{SW} - G_y^W$ are equal, they are 0.005% of *RE*, which means 1.5% of the maximum value attained by R^{EG} (bandwidth

10 euro large) and 1% of the value attained by $\left(G_{y-T}^{SW}-G_{y}^{W}\right)$ at 3000 euro bandwidth; this explains why at a 280 euro bandwidth the three potential vertical redistribution indexes look to be equal. Slightly larger percentages hold for family equivalent incomes [20]:

We can notice that the maximum for V^{AJL}, the lowest among the three indexes, lies quite close to the point where V^{VCL} crosses V. This implies that the bandwidth where $R^{EG}=G_{y-T}^{SW}-G_{y}^{W}$ is in a neighborhood of bandwidths where V^{AJL} is maximum. Observe that $V^{AJL}\le V^{VCL}$ holds together with $V^{AJL}\le V$, so that V^{VCL} crosses V when the former is still increasing and the latter is already decreasing. If this is true, the global maximum for V^{AJL} should fall, as it actually does, between the bandwidth where the separation between V^{VCL} and V^{AJL} becomes evident, and the bandwidth where V^{VCL} and V becomes no longer distinguishable.

All these considerations seam to show that the real issue is to obtain a potential vertical measure which should be similar for each of the three indexes, without foregoing the VCL idea that this measure should be as great as possible.

Then we could state the following criterion: the optimal bandwidth should be identified as follows:

$$\min\left\{\frac{\max\left[\left|V^{VCL}-V\right|;\left|V^{AJL}-V\right|;\left|V^{VCL}-V^{AJL}\right|\right]}{\min\left[V^{VCL};V^{AJL};V\right]}\right\} \tag{9}$$

[20] R^{EG} and $\left(G_{y-T}^{SW}-G_{y}^{W}\right)$ represent 0.012% of RE at the 400 euro bandwidth; which means 2.3% of the maximum value attained by R^{EG} (bandwidth 10 euro large) and 2.1% of the maximum attained by $\left(G_{y-T}^{SW}-G_{y}^{W}\right)$ (bandwidth 3000 euro).

The criterion stated at (9) should avoid choosing bandwidths where measures of the potential redistributive effects are lower than *RE* or quite null[21].

Figures 4a and 4b, 5a, 5b ABOUT HERE

Figures 5a and 5b show the behavior of criterion (9) for the bandwidth range 10-3000. As we can see, the lines of (9) are quite stable if compared to the upper traces of R^{EG} and $\left(G_{y-T}^{SW} - G_{y}^{W}\right)$, and they appear just as a re-scaling of the two distances.

By observing the figures, we can conclude that the minimum for (9) is reached when $R^{EG} = \left(G_{y-T}^{SW} - G_{y}^{W}\right)$, at a 280 euro bandwidth for individuals and at a 380 euro bandwidth for equivalent family incomes.

We expect that (9) should further increase as bandwidth enlarges; moreover, for bandwidths larger than those considered in this article, (9) can be approximated by $\left(V^{VCL} - V^{AJL}\right)/V^{AJL} = \left(V^{VCL}/V^{AJL}\right) - 1 = \left\{\left(G_{y}^{B} - G_{y-T}^{B}\right)/\left[\left(G_{y}^{B} - G_{y-T}^{B}\right) - \left(G_{y-T}^{SW} - G_{y}^{W}\right)\right]\right\} - 1$.

For bandwidths up to 3000 euro, $\left(G_{y-T}^{SW} - G_{y}^{W}\right)$ is quite small with respect to $\left(G_{y}^{B} - G_{y-T}^{B}\right)$, so that $\left(V^{VCL}/V^{AJL}\right)$ is quite close to 1; however, when bandwidths become larger and larger, $\left(G_{y-T}^{SW} - G_{y}^{W}\right)$ becomes relatively more important with respect to $\left(G_{y}^{B} - G_{y-T}^{B}\right)$, which causes $\left(V^{VCL}/V^{AJL}\right)$ to assume values sensibly greater than 1.

[21] Mussini (2008) intensively investigates the behaviour of criterion (9) for the income earners in the Municipality of Milano. He explores for bandwidths up to ½ the maximum range for, confirming that the criterion (9) gives the minimum value where $R^{EG} = \left(G_{y-T}^{SW} - G_{y}^{W}\right)$.

We add that in our empirical analysis, in the interval between 250-370 euro, V^{AJL} oscillates between 99.97% and 99.99% of the Reynolds-Smolensky index, for individuals; for families the percentage ranges from 99.97% to 99.98% when the interval is 340-440 euro: in a neighbourhood of the optimal bandwidth the three indexes absorb most of the total redistributive effect. We conclude by observing that at the optimal bandwidth the horizontal loss measured by H^{VCL} and H^{AJL} is much lower than the loss due to overlapping among groups, measured by R^{AJL}.

6. Conclusions

The original Aronson, Johnson and Lambert (1994) decomposition of the redistributive effect considers groups of exact equals in portioning the whole pre-tax income distribution and restricts the analysis to the special cases in which the group averages and the within group orderings maintain the same ranking as before taxation. This means that the AJL decomposition of the redistributive effect considers only the overlapping effect among groups of exact equals.

As the following literature points out, exact equals are rare in real world data, so that only groups with close pre-tax incomes can be considered. If this is the case, also the reranking of the mean post-tax income among groups and the reranking within groups must be considered. The intensity of the three possible rerankings considered here varies according to the bandwidth defining the close equals. Then a problem arises: an optimal bandwidth must be chosen in order to decompose the redistributive effect properly into vertical, horizontal and reranking effects.

The choice of the optimal bandwidth is not obvious. Van de Ven, Creedy and Lambert (2001) identify the optimal bandwidth that should be used in decomposing the redistributive effect as the Aronson, Johnson and Lambert (1994) methodology suggests, without considering the different contribution of the reranking of the mean post-tax income among groups and the reranking within groups. They suggest choosing as the optimal bandwidth the one that maximizes the ratio between the potential vertical effect and the actual redistributive effect. As the empirical analysis shows, this ratio may have more than one relative maximum and present a layout that may be irregular, so that this condition is difficult to be applied in real data elaborations.

Urban and Lambert (2008) solved this problem by identifying a set of possible decompositions of the redistributive effect. They also notice that when close pre-tax equals groups instead of exact pre-tax ones are considered, the residual component in the original Aronson, Johnson and Lambert (1994) model is not the Atkinson-Plotnick-Kakwani index, but only one of its components, that is the one which measures group overlapping introduced by taxation.

In this paper we have used this decomposition of the Atkinson-Plotnick-Kakwani index, and intensively looked at the empirical analysis in order to identify the relationships between the three main possible decompositions of the redistributive effect analyzed by Urban and Lambert (2008).

We conclude that the optimal bandwidth should be chosen where the ratio between (a) the maximum distance among the considered possible definitions of potential vertical effect and (b) the minimum among the potential vertical effects is minimum. We find empirical evidence that in this bandwidth neighborhood the three measures also very nearly converge and, moreover, absorb most of the Reynolds-Smolenky total redistribution measure.

References

Aronson R. J., Lambert P. J., (1993), "Inequality decomposition analysis and the Gini coefficient revisited", *The Economic Journal*, 103, pp. 1221-1227.

Aronson R. J., Lambert. P. J., (1994), "Decomposing the Gini coefficient to reveal the vertical, horizontal and reranking effects of income taxation", *National Tax Journal*, 47, pp. 273-294.

Aronson R. J., Johnson P. J., Lambert P. J., (1994), "Redistributive effect and unequal income tax treatment", *The Economic Journal*, 104, pp. 262-270.

Dagum C. (1997), "A new approach to the decomposition of Gini income inequality ratio", *Empirical Economics*, 22, pp. 515-531.

Ebert U., Moyes P., (2000), "Consistent income tax structures when households are heterogeneous", *Journal of Economic Theory*, 90, pp. 116-150.

Lambert P.J., Ramos X., (1997a), "Horizontal inequity and vertical redistribution", *International Tax and Public Finance*, 4, pp. 25-37.

Lambert P.J., Ramos X., (1997b), "Horizontal inequity and reranking", *Research on Economic Inequality*, 7, pp. 1-18.

Mussini M., (2008), "A note on choosing a proper income class width when decomposing redistributive effect", Università di Milano Bicocca, Dipartimento di Statistica, Quaderno di Dipartimento n. 20.

Pellegrino S., (2007a), "Struttura ed effetti redistributivi dell'imposta personale e italiana: il confronto 2000-2005 ed un esercizio di modifica", *Economia Pubblica*, 37, n. 1-2, pp. 99-143.

Pellegrino S., (2007b), "Il Modello di microsimulazione IRPEF 2004", *SIEP*, *Italian Society of Public Economics*, WP 583/07.

Pellegrino S., (2007c), "IRPEF 2007: una redistribuzione (quasi) irrilevante?", *Rivista di Diritto Finanziario e Scienza delle Finanze*, anno LXVI, fasc. 1, 2007, pp. 24-43.

Urban I., Lambert P. J., (2008), "Redistribution, horizontal inequity and reranking: how to measure them properly", *Public Finance Review*, 20, n. 10, pp.1-24.

van de Ven J., Creedy J., Lambert P. J., (2001), "Close equals and calculation of the vertical, horizontal and reranking effects of taxation", *Oxford Bulletin of Economics and Statistics*, 63, pp. 381-394.

Vernizzi A., (2007), "Una precisazione sulla scomposizione dell'indice di redistribuzione *RE* di Aronson-Johnson-Lambert e una proposta di estensione dell'indice di Plotnick", *Economia Pubblica*, 37, n. 1-2, pp. 145-153, and *DEAS*, *Università degli Studi di Milano*, WP 2006-28.

Vernizzi A., Pellegrino S. (2007), "On the Aronson-Johnson-Lambert decomposition of the redistributive effect", *DEAS*, *Università degli Studi di Milano*, WP 2007-13.

Table 1: Summary of index definitions

Groups	are constituted by subjects belonging to a same pre-tax income bracket; income brackets are created by splitting the pre-tax non decreasing incomes parade into contiguous intervals characterized by a same income spread. Groups contains the same subjects both before and after taxation, whatever ordering criterion is adopted. Before taxation no overlapping exists by construction; taxation may result in group overlapping.
G_y	Gini index for pre-tax income parade.
G_y^B	between groups Gini index for pre-tax income parade: it is defined as the Gini index when all incomes inside each group are substituted by the group income average.
G_y^W	within groups Gini index for pre-tax income parade: $G_y^W = \sum_k a_{k,y} G_{k,y}$, where $G_{k,y}$ is the Gini index for the *k*-th group and $a_{k,y}$ is the product of the *k*-th group population share and pre-tax income share.
G_{y-T}	Gini index for post-tax income parade.
G_{y-T}^B	it is analogous to G_y^B for the post-tax income parade.
G_{y-T}^W	within groups Gini index for post-tax income parade: $G_{y-T}^W = \sum_k a_{k,y-T} G_{k,y-T}$, where $G_{k,y-T}$ is the post-tax Gini index for the *k*-th group and $a_{k,y-T}$ is the product of the *k*-th group population share and post-tax income share.
D_{y-T}	concentration index for post-tax income parade when ordered according to the pre-tax order.
D_{y-T}^B	between groups concentration index for post-tax income parade: it is defined as the concentration index when all incomes inside each group are substituted by the group income average, moreover groups are ordered according to pre-tax group averages.
D_{y-T}^W	within groups concentration index for post-tax income parade: $D_{y-T}^W = \sum_k a_{k,y-T} D_{k,y-T}$; $D_{k,y-T}$ is the concentration index for the *k*-th group, when the *k*-th group incomes are ordered according to the pretax within group order, and $a_{k,y-T}$ is the product of the *k*-th group population share and post-tax income share.
G_{y-T}^{SW}	within groups Gini index for post-tax smoothed income parade. Smoothed taxation consists in taxing all income in a group by the group average tax rate. $G_{y-T}^{SW} = \sum_k a_{k,y-T} G_{k,y-T}$, as the Gini index for the *k*-th group remains unchanged, when all group incomes are taxed by a same tax rate.

Table 2: Summary of equations and components

$RE = V^{VCL} - H^{VCL} - R^{AJL}$ $V^{VCL} = G_y^B - G_{y-T}^B$ $H^{VCL} = G_{y-T}^W - G_y^W$ $R^{AJL} = G_{y-T}^t = G_{y-T} - G_{y-T}^B - G_{y-T}^W$ $\lim_{b\to 0} V^{VCL} = RE \quad \lim_{b\to 0} H^{VCL} = 0 \quad \text{e} \quad \lim_{b\to 0} R^{AJL} = 0$ $\lim_{b\to MAX} V^{VCL} = 0 \quad \lim_{b\to MAX} H^{VCL} = -RE \quad \text{e} \quad \lim_{b\to MAX} R^{AJL} = 0$
$RE = V^{AJL} - H^{AJL} - R^{AJL}$ $V^{AJL} = G_y - G_{y-T}^B - G_{y-T}^{SW} =$ $= V^{VCL} - \left(G_{y-T}^{SW} - G_y^W\right)$ $H^{AJL} = G_{y-T}^W - G_{y-T}^{SW}$ $\lim_{b\to 0} V^{AJL} = RE \quad \lim_{b\to 0} H^{AJL} = 0 \quad \text{e} \quad \lim_{b\to 0} R^{AJL} = 0$ $\lim_{b\to MAX} V^{AJL} = 0 \quad \lim_{b\to MAX} H^{AJL} = -RE \quad \text{e} \quad \lim_{b\to MAX} R^{AJL} = 0$
$RE = V - H - R^{APK}$ $V = G_y - D_{y-T}^B - G_{y-T}^{SW} =$ $= V^{AJL} + \left(G_{y-T}^B - D_{y-T}^B\right) =$ $= V^{VCL} - \left(G_{y-T}^{SW} - G_y^W\right) + \left(G_{y-T}^B - D_{y-T}^B\right)$ $H = D_{y-T}^W - G_{y-T}^{SW}$ $\lim_{b\to 0} V = G_y - D_{y-T} \quad \text{e} \quad \lim_{b\to 0} H = 0$ $\lim_{b\to MAX} V = 0 \quad \text{e} \quad \lim_{b\to MAX} H = D_{y-T} - G_y$
$R^{APK} = R^{AJL} + R^{EG} + R^{WG}$ $R^{EG} = \left(G_{y-T}^B - D_{y-T}^B\right)$ $R^{WG} = \left(G_{y-T}^W - D_{y-T}^W\right)$ $\lim_{b\to 0} R^{AJL} = 0 \quad \lim_{b\to 0} R^{EG} = R^{APK} \quad \text{e} \quad \lim_{b\to 0} R^{WG} = 0$ $\lim_{b\to MAX} R^{AJL} = 0 \quad \lim_{b\to MAX} R^{EG} = 0 \quad \text{e} \quad \lim_{b\to MAX} R^{WG} = R^{APK}$

Figure 1a: V, V^{VCL} and V^{AJL} (%*RE*) - Individuals

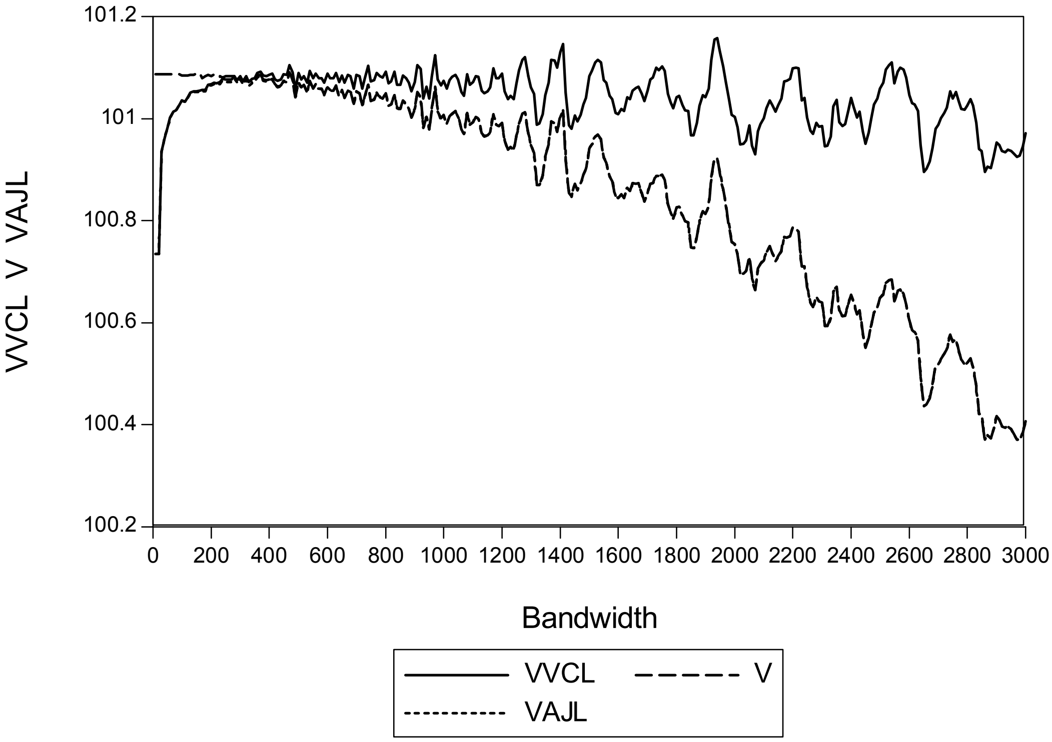

Figure 1a bis: V, V^{VCL} and V^{AJL} (%*RE*) – Individuals (focus)

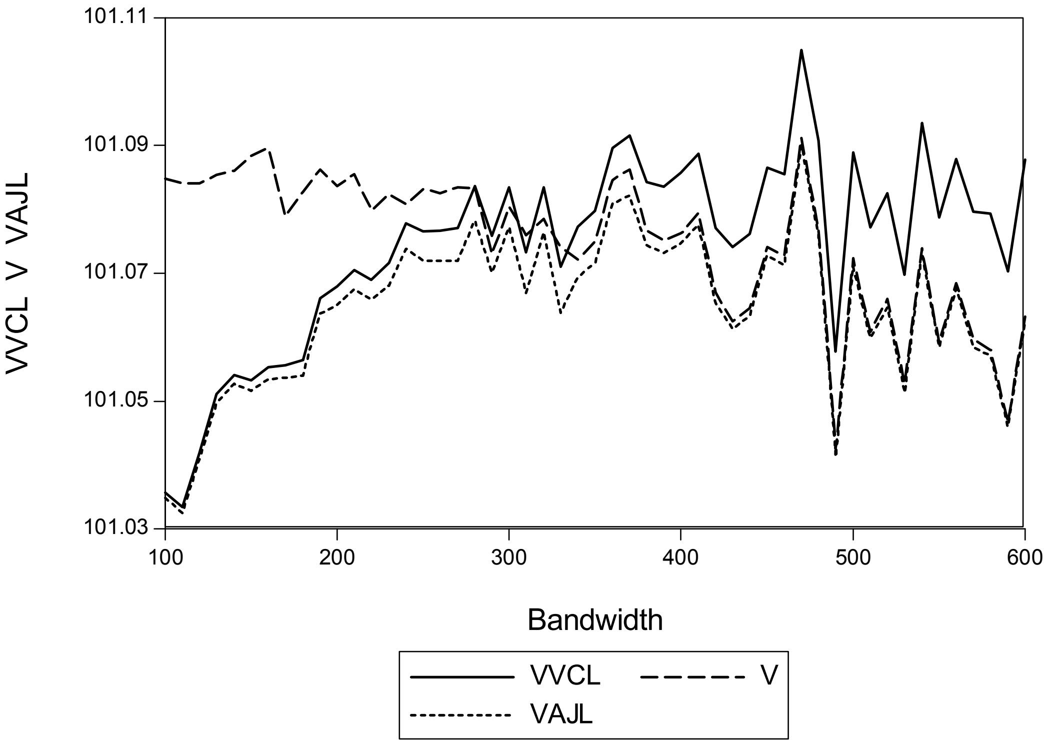

Figure 1b: V, $V^{VCL}$and V^{AJL} (%RE) – m(1) - Households

m(1) means Cutler scale α=0.50 β=0.65 and family weight = 1

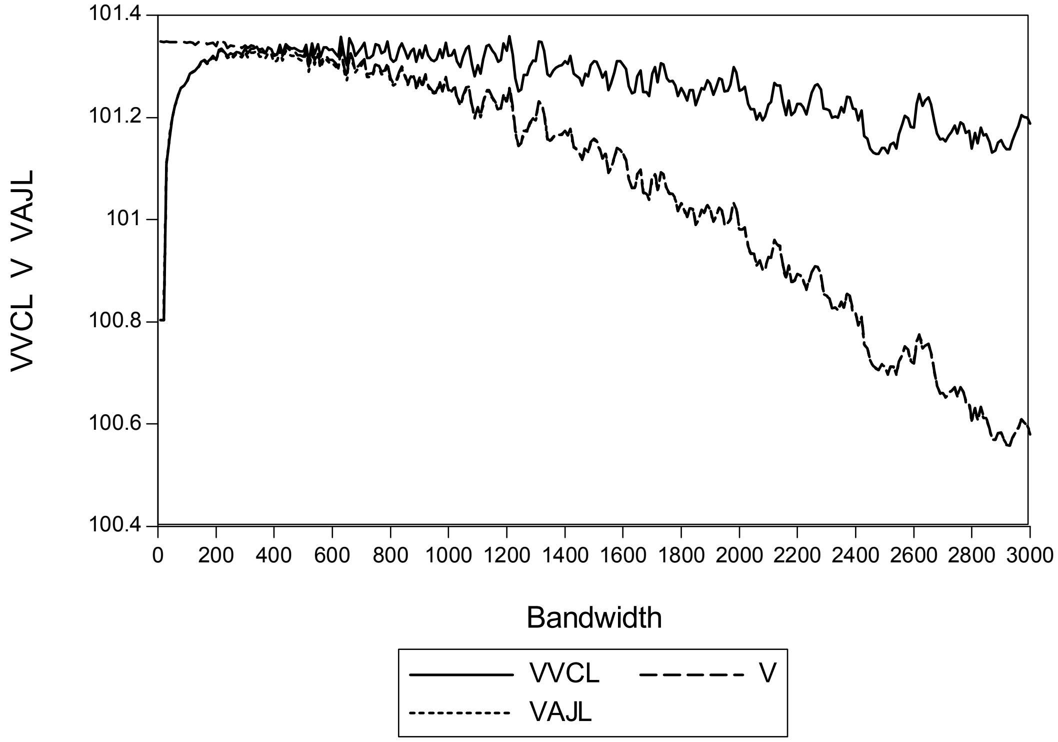

Figure 1b bis: V, $V^{VCL}$and V^{AJL} (%RE) – m(1) - Households (focus)

m(1) means Cutler scale α=0.50 β=0.65 and family weight = 1

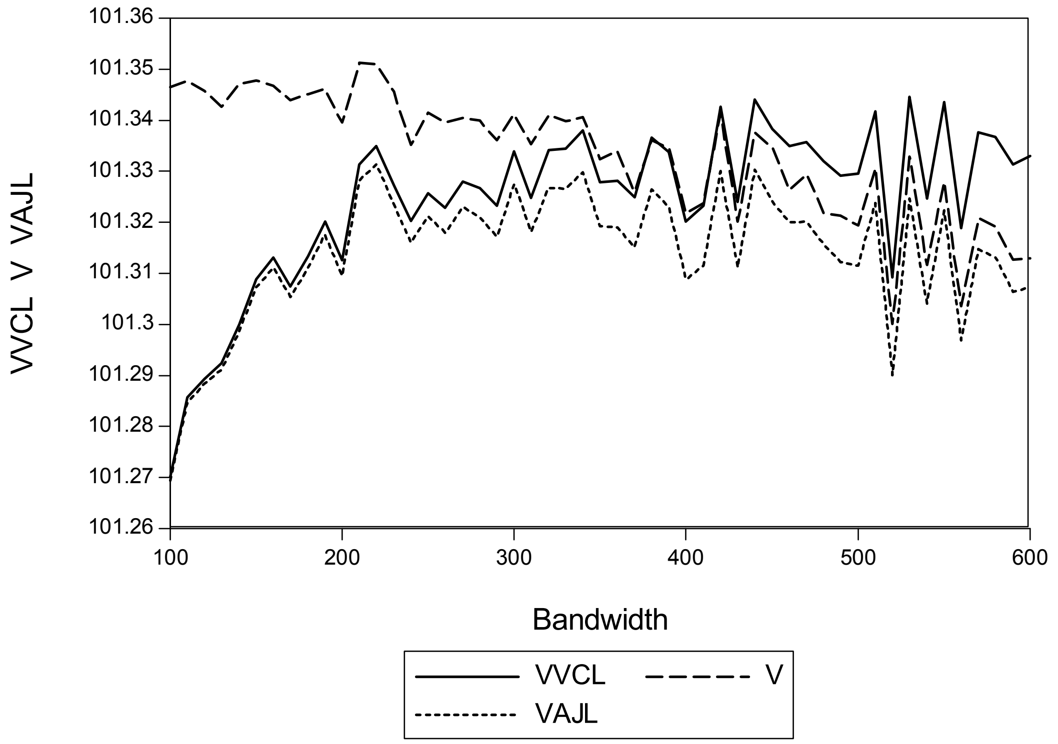

Figure 2a: H, H^{VCL} and H^{AJL} with R^{AJL} and R^{APK} (%RE) - Individuals

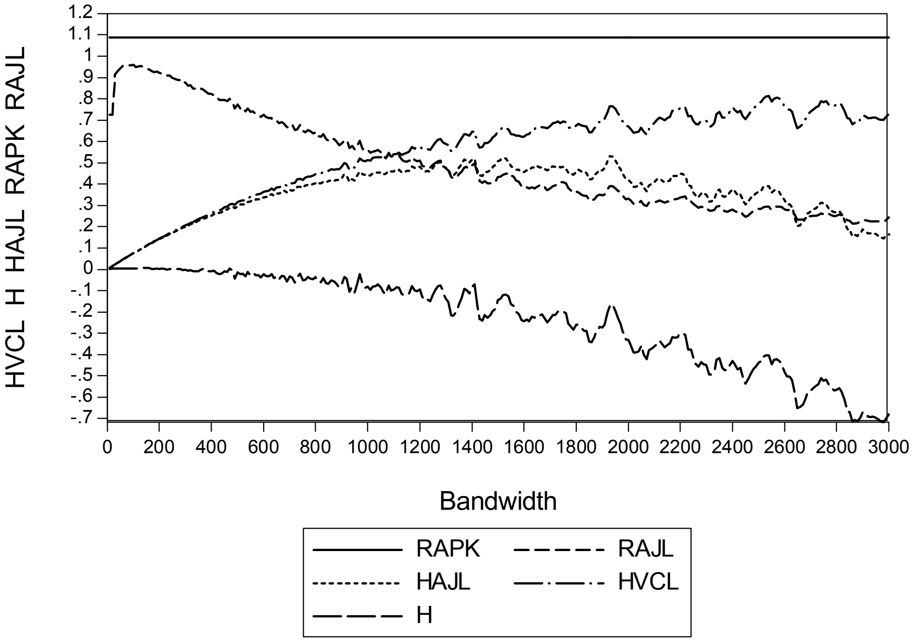

Figure 2b: H, H^{VCL} and H^{AJL} with R^{AJL} and R^{APK} (%RE) - m(1) - Households

§ m(1) means Cutler scale α=0.50 β=0.65 and family weight = 1.

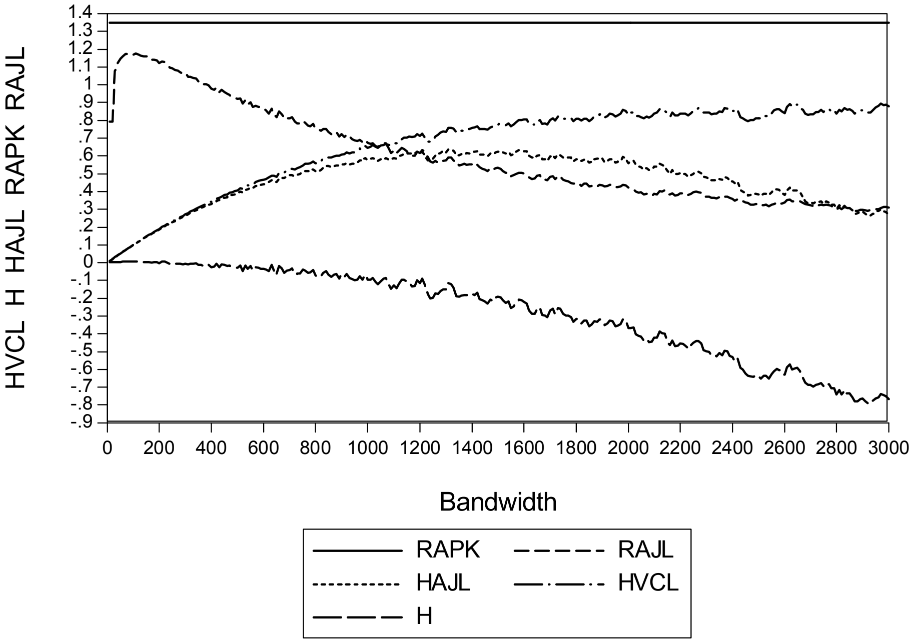

Figure 3a: R^{APK} % decomposition - Individuals

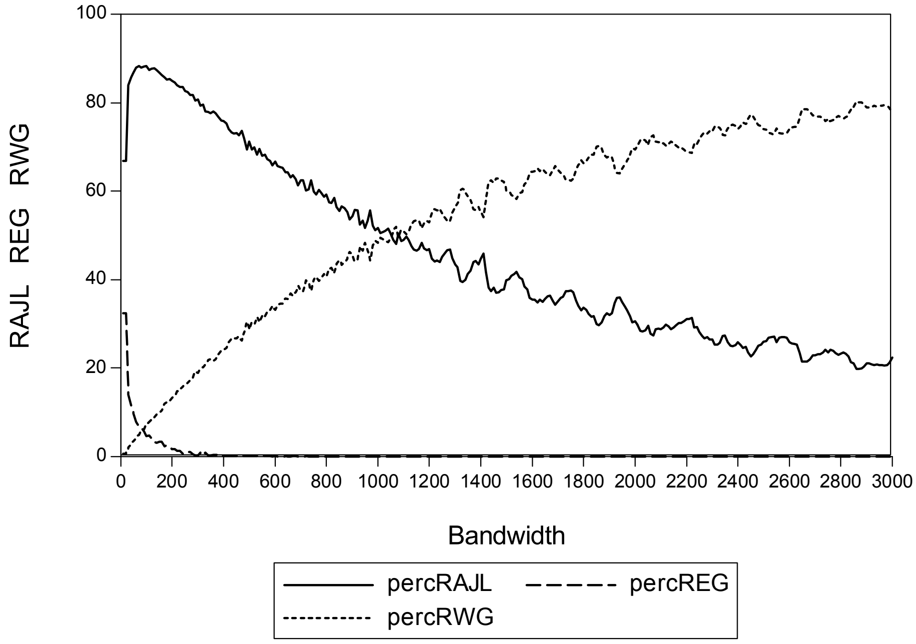

Figure 3b: R^{APK} % decomposition – $m(1)$ - Households

§ $m(1)$ means Cutler scale α=0.50 β=0.65 and family weight = 1.

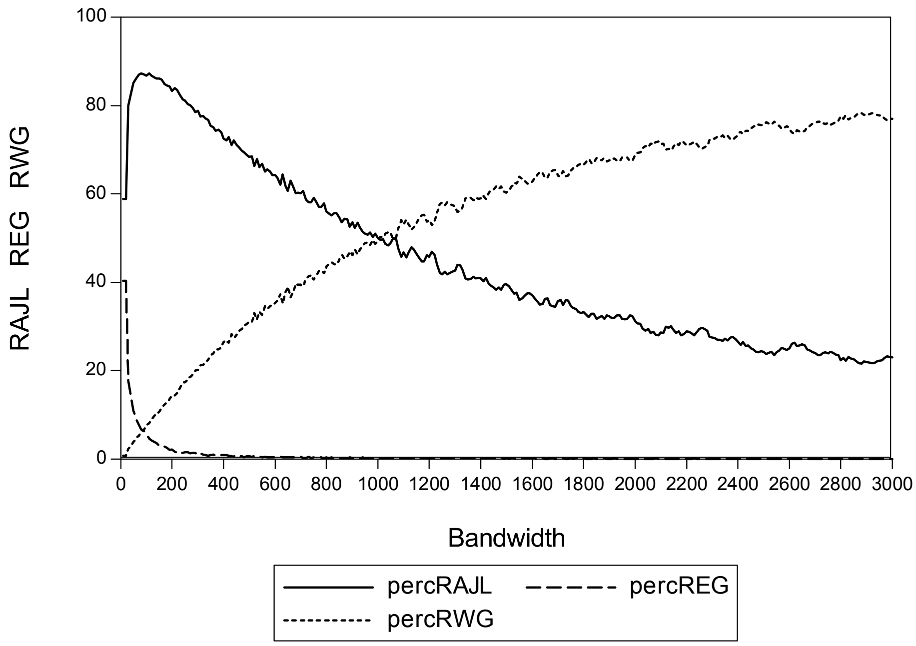

Table 3: *RE* decomposition – Individuals

(bootstrap estimated standard errors in parentheses-2,000 replications)

Component	*Bandwidths*						
	100	300	400	500	600	700	2000
% RE/G_y	14.3699 (0.1266)	14.3699 (0.1266)	14.3699 (0.1266)	14.3699 (0.1266)	14.3699 (0.1266)	14.3699 (0.1266)	14.3699 (0.1266)
$\%\left(V^{VCL}/RE\right)$	101.0357 (0.0381)	101.0835 (0.0388)	101.0858 (0.0390)	101.0890 (0.0392)	101.0878 (0.0397)	101.0855 (0.0392)	101.0039 (0.0380)
$\%\left(V/RE\right)$	101.0847 (0.0395)	101.0806 (0.0387)	101.0762 (0.0392)	101.0723 (0.0392)	101.0634 (0.0398)	101.0537 (0.0390)	100.7546 (0.0375)
$\%\left(H^{VCL}/RE\right)$	0.0759 (0.0025)	0.2057 (0.0064)	0.2621 (0.0083)	0.3157 (0.0098)	0.3628 (0.0115)	0.4065 (0.0126)	0.6713 (0.0232)
$\%\left(H/RE\right)$	0.0022 (0.0012)	0.0063 (0.0031)	0.0107 (0.0042)	0.0146 (0.0052)	0.0235 (0.0061)	0.0332 (0.0070)	0.3323 (0.0222)
$\%\left(R^{AJL}/RE\right)$	0.9598 (0.0349)	0.8778 (0.0334)	0.8237 (0.0316)	0.7733 (0.0307)	0.7250 (0.0283)	0.6789 (0.0270)	0.3325 (0.0170)
$\%\left(R^{APK}/RE\right)$	1.0871 (0.0391)	1.0871 (0.0391)	1.0871 (0.0391)	1.0871 (0.0391)	1.0871 (0.0391)	1.0871 (0.0391)	1.0871 (0.0391)
$\%\left(R^{AJL}/R^{APK}\right)$	88.2865 (0.5885)	80.7452 (0.4274)	75.7697 (0.4252)	71.1290 (0.4923)	66.6890 (0.5651)	62.4498 (0.5982)	30.5890 (0.8749)
$\%\left(R^{EG}/R^{APK}\right)$	4.5962 (0.6003)	0.2900 (0.3100)	0.1338 (0.1310)	0.1116 (0.1115)	0.0669 (0.0549)	0.0669 (0.0495)	0.0000 (0.0146)
$\%\left(R^{WG}/R^{APK}\right)$	7.1174 (0.1467)	18.9648 (0.3454)	24.0964 (0.4105)	28.7595 (0.5073)	33.2441 (0.5693)	37.4833 (0.6197)	69.4110 (0.9081)

Source: Own elaborations.

Table 4: *RE* decomposition – Households

(bootstrap estimated standard errors in parentheses-2,000 replications)

Component	*Bandwidths*						
	100	300	400	500	600	700	2000
% RE/G_y	13.9266 (0.1910)	13.9266 (0.1910)	13.9266 (0.1910)	13.9266 (0.1910)	13.9266 (0.1910)	13.9266 (0.1910)	13.9266 (0.1910)
$\%\left(V^{VCL}/RE\right)$	101.2702 (0.0543)	101.3349 (0.0574)	101.3201 (0.0577)	101.3300 (0.0574)	101.3330 (0.0575)	101.3351 (0.0565)	101.2527 (0.0565)
$\%\left(V/RE\right)$	101.3468 (0.0606)	101.3412 (0.0577)	101.3217 (0.0574)	101.3193 (0.0585)	101.3129 (0.0592)	101.3079 (0.0592)	100.9817 (0.0574)
$\%\left(H^{VCL}/RE\right)$	0.1001 (0.0040)	0.2718 (0.0106)	0.3422 (0.0136)	0.4078 (0.0158)	0.4664 (0.0188)	0.5245 (0.0202)	0.8316 (0.0335)
$\%\left(H/RE\right)$	0.0016 (0.0017)	0.0072 (0.0044)	0.0266 (0.0061)	0.0290 (0.0074)	0.0354 (0.0091)	0.0405 (0.0101)	0.3667 (0.0313)
$\%\left(R^{AJL}/RE\right)$	1.1701 (0.0504)	1.0622 (0.0486)	0.9780 (0.0453)	0.9217 (0.0433)	0.8666 (0.0420)	0.8106 (0.0394)	0.4210 (0.0246)
$\%\left(R^{APK}/RE\right)$	1.3486 (0.0605)	1.3486 (0.0605)	1.3486 (0.0605)	1.3486 (0.0605)	1.3486 (0.0605)	1.3486 (0.0605)	1.3486 (0.0605)
$\%\left(R^{AJL}/R^{APK}\right)$	86.7615 (0.7555)	78.7591 (0.5072)	72.5153 (0.5598)	68.3462 (0.6584)	64.2561 (0.6792)	60.1067 (0.7271)	31.2191 (1.0030)
$\%\left(R^{EG}/R^{APK}\right)$	5.7103 (0.8019)	1.0275 (0.3685)	0.9484 (0.3194)	0.5928 (0.2602)	0.4347 (0.1983)	0.4742 (0.2152)	0.0395 (0.0384)
$\%\left(R^{WG}/R^{APK}\right)$	7.5282 (0.1758)	20.2134 (0.4143)	26.5363 (0.5429)	31.0611 (0.6269)	35.3092 (0.6628)	39.4191 (0.7425)	68.7414 (0.9690)

Source: Own elaborations.

Figure 4a: R^{EG} and $G^{SW}_{y-T} - G^{W}_{y}$ as percentage of RE – Individuals

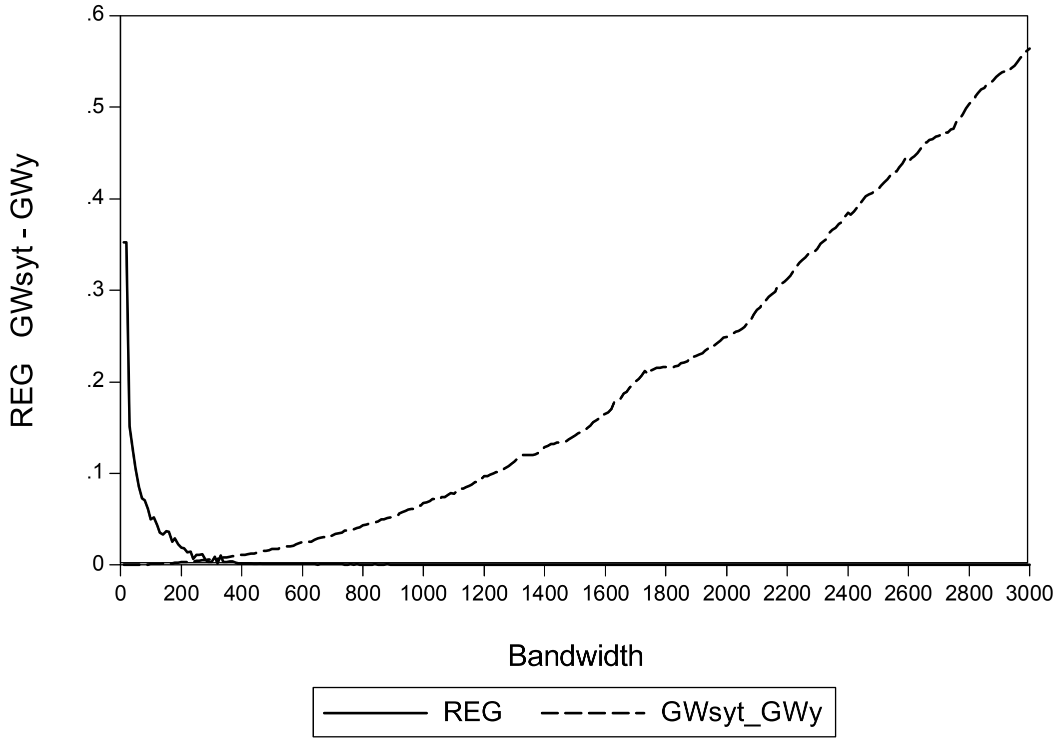

Figure 4b: R^{EG} and $G^{SW}_{y-T} - G^{W}_{y}$ as percentage of RE – *m*(1) - Households

§ *m*(1) means Cutler scale α=0.50 β=0.65 and family weight = 1.

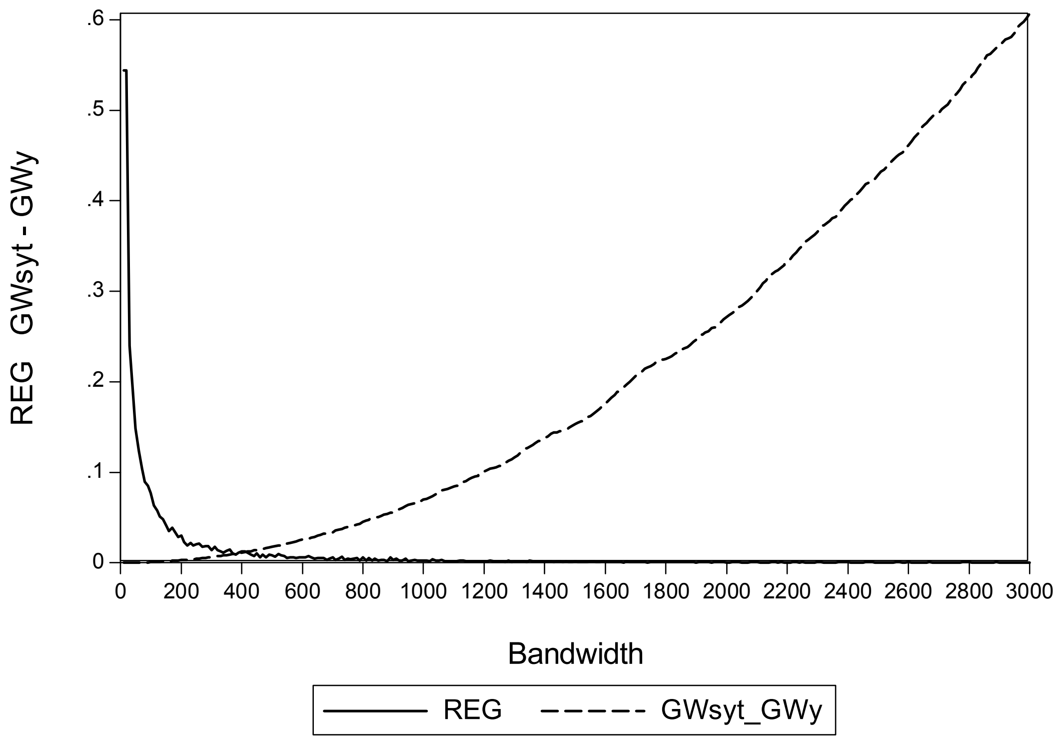

Figure 5a: The minimization criterion – Individuals

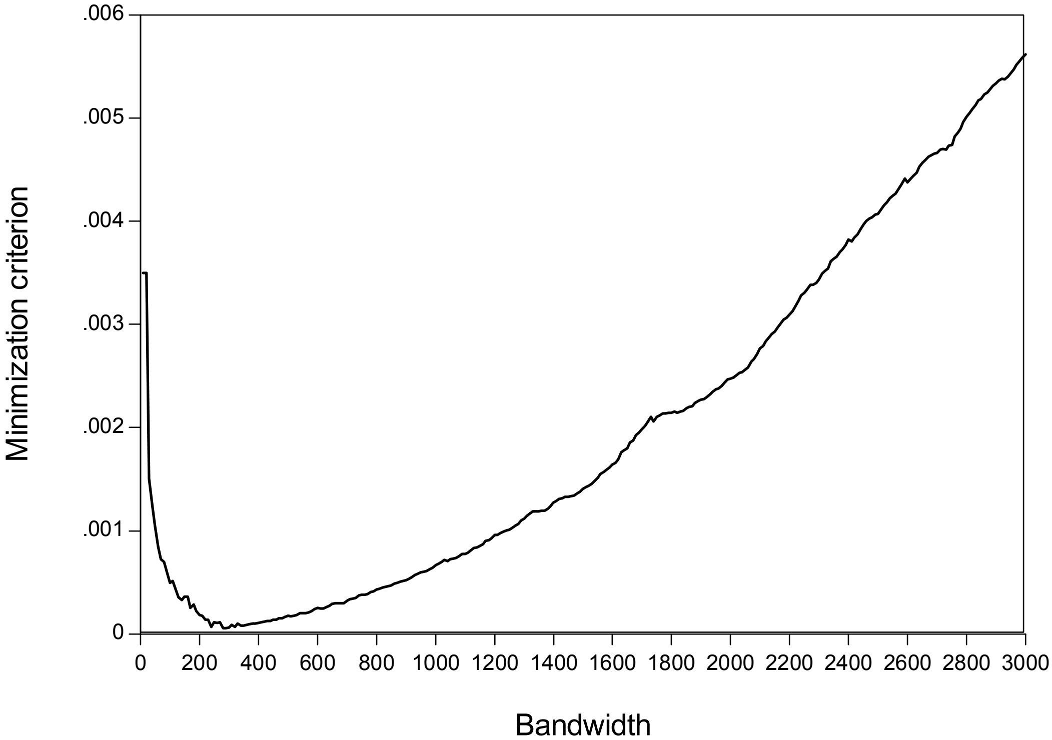

Figure 5b: The minimization criterion – *m*(1) - Households