The Duclos-Jalbert-Araar decomposition of redistributive effect: implementation issues

Ivica, Urban

Institute of Public Finance

11 January 2011

The Duclos-Jalbert-Araar Decomposition of Redistributive Effect: Implementation Issues

IVICA URBAN

Institute of Public Finance, Zagreb, Croatia, ivica.urban@ijf.hr

Version: May 5, 2012

Abstract

Models decomposing the redistributive effect of fiscal systems into vertical and horizontal effects are extensively used by practitioners. The Duclos, Jalbert, and Araar's (2003) model, despite its advantages, has not yet been widely employed in empirical research, possibly due to a relatively challenging implementation procedure that involves the estimation of expected post-fiscal incomes. To override these difficulties the designers of the software DAD (Duclos, Araar, and Fortin 2010) have incorporated a module for implementation of the model. However, the application of this module on Croatian tax-benefit system data revealed certain inaccuracies in the results. Carefully unfolding the calculation and estimation procedures needed for implementation of the model, this paper instructs practitioners on how to correctly apply the model and helps DAD designers improve their module.

1 INTRODUCTION

Duclos, Jalbert, and Araar (2003) (henceforth DJA) have designed a comprehensive model to decompose the redistributive effect (RE) of a fiscal system into vertical, classical horizontal inequity (henceforth CHI), and reranking effects. The model is built into the framework of the Atkinson-Gini social welfare function (henceforth AGF), which first converts incomes into utilities employing the Atkinson's (1970) utility function, and then aggregates utilities using rank-dependent weights, which underlie the S-Gini and S-concentration coefficients proposed by Donaldson and Weymark (1980) and Yitzhaki (1983).[1]

The DJA model has certain advantages over its competitors, the widely acknowledged Kakwani's (1984) (henceforth K84) and the Aronson, Johnson, and Lambert's (1994) (henceforth AJL) decompositions of RE. To measure the CHI effect, the researcher must determine the set of counterfactual CHI-free or expected post-fiscal incomes (EPIs). While the AJL model relies on the formation of arbitrary groups of close equals in this task, the DJA model employs purposefully designed statistical procedures. Consequently, the implementation of the DJA model requires a certain expertise related to data smoothing and curve-fitting methods. To facilitate the application of the DJA model in empirical analysis, a module for calculation of the DJA indices from the sample data is incorporated into the software DAD (Duclos, Araar, and Fortin 2010) (henceforth DAD-DJA).

The use of DAD-DJA in research on the Croatian tax-benefit system revealed certain inaccuracies in the results. Specifically, when the ethical parameter of AGF is set to zero, the CHI effect in the DJA model should be equal to zero by construction. However, the estimated value of the CHI effect was significantly different from zero. Analysis has shown that DAD-DJA produces upward biased estimates of EPIs in the low pre-fiscal income region. Furthermore, it was revealed that the fitting procedure in DAD-DJA contains a 'bug', producing unreasonably high estimates of EPIs for the top pre-fiscal income units in the sample.

In an attempt to obtain fully accurate estimates of DJA indices, independent procedures have been developed. They are thoroughly explained in this paper to assist practitioners in implementing the DJA model and to help DAD designers improve the working of DAD-DJA. A brief overview of data smoothing methods is provided, accompanied by advice on how to accurately obtain EPIs estimates. Relationships with other measurement models are explained.

The rest of the paper is organized as follows. Section 2 briefly exposes the elements of the DJA model and its connections with other decompositions. Section 3 extensively describes the procedures of data preparation, estimation and calculation of various elements of the DJA model, and employs them on a simple hypothetical population of four income units. In section 4 the procedures are applied to data on the Croatian tax-benefit system, and the results are compared with those obtained by DAD-DJA. Section 5 concludes the paper.

2 THE DUCLOS-JALBERT-ARAAR MODEL

Post-fiscal income is equal to pre-fiscal income minus taxes plus benefits. The change of income inequality induced by a fiscal system consisting of taxes and benefits is called the redistributive effect (RE). In measurement terms, we have that $\Delta = I(X) - I(N)$, where Δ represents RE, and $I(X)$ and $I(N)$ are indices of pre- and post-fiscal income inequality.

In the DJA model, inequality indices $I(\cdot)$ are derived using the Atkinson-Gini social welfare function, proposed by Araar and Duclos (2003, 2006). For pre-fiscal income we have that

$$W(X, \varepsilon, v) = \int_0^1 U(X(p), \varepsilon)\, \omega(p, v) dp \tag{1}$$

where ε is the ethical parameter configuring the Atkinson's (1970) utility function, $U(X(p), \varepsilon) = [X(p)]^{1-\varepsilon}/(1-\varepsilon)$ for $\varepsilon \neq 1$, and $U(X(p), \varepsilon) = \ln[X(p)]$ for $\varepsilon = 1$, with p denoting the quantiles of the pre-fiscal income distribution, and $X(p)$ the income at quantile p. The term v is another ethical parameter, characterizing the Donaldson and Weymark's (1980) and Yitzhaki's (1983) S-Gini rank-dependent weighting scheme, $\omega(p, v) = v(1-p)^{v-1}$. The equally distributed equivalent income is

an inverse function of $W(\cdot)$ and is obtained as $\xi(X, \varepsilon, \nu) = [(1 - \varepsilon)W(X, \varepsilon, \nu)]^{1/(1-\varepsilon)}$ for $\varepsilon \neq 1$, and

$\xi(X, \varepsilon, \nu) = \exp[W(X, \varepsilon, \nu)]$ for $\varepsilon = 1$. Finally, the Atkinson-Gini inequality index is calculated as

follows:

$$I(X) = 1 - \xi(X, \varepsilon, \nu)/\mu^X \tag{2}$$

where μ^X is the mean pre-fiscal income. $I(N)$ is obtained analogously, using the quantiles of the

post-fiscal income distribution.

The DJA model decomposes RE as follows:

$$\Delta = V - C - R = [I(X) - I(N^E)] - [I(U^P) - I(N^E)] - [I(N) - I(U^P)] \tag{3}$$

The vertical effect, $V = I(X) - I(N^E)$, represents the potential RE or the reduction of

inequality that would be achieved by the counterfactual, CHI-free system. The discrepancy between

potential and actual RE is divided into a CHI effect, $C = I(U^P) - I(N^E)$, and a reranking effect,

$R = I(N) - I(U^P)$, which measure two different manifestations of horizontal inequity (HI). The

former effect (C) measures HI emerging from violation of the 'classical horizontal equity principle',

which says that equals should be treated equally. The latter effect (R) evaluates HI arising from

infringement of the 'no-reranking principle', which requires the fiscal process to not change the

ranks of income units in transition from pre- to post-fiscal income. Take, for example, four

households of equal size. A and B have pre-fiscal incomes of 10 each, whereas C and D have pre-

fiscal incomes of 20 each. Suppose that A, B, C, and D end up with post-fiscal incomes of 8, 16, 12,

and 24, respectively. Between pre-fiscal equals (A and B; C and D) CHI has occurred, whereas

between pre-fiscal unequals (B and C) reranking has taken place.

In equation (3), $I(N^E)$ is the inequality index obtained for EPIs, equal to $N^E(p) = \int_0^1 N(q|p)dq$, where $N(q|p)$ denotes a post-fiscal income at the qth quantile among all those income

units belonging to the pth quantile of the pre-fiscal income distribution. $I(U^P)$ represents the

inequality index obtained for expected post-fiscal utilities (EPUs) at quantile p, $U^P(p, \varepsilon) = \int_0^1 U(N(q|p), \varepsilon) dq$. For $N^E(p)$ and $U^P(p, \varepsilon)$ we obtain the respective social welfare functions

$W(N^E, \varepsilon, \nu) = \int_0^1 U(N^E(p), \varepsilon) \, \omega(p, \nu) dp$ and $W(U^P, \varepsilon, \nu) = \int_0^1 U^P(p, \varepsilon) \, \omega(p, \nu) dp$, while the

corresponding inequality indices are $I(N^E) = 1 - \xi(N^E, \varepsilon, \nu)/\mu^N$ and $I(U^P) = 1 - \xi(U^P, \varepsilon, \nu)/\mu^N$,

where μ^N is the mean post-fiscal income.

In the special case where $\varepsilon = 0$, utilities are identical to incomes: $U(y, 0) = y$. Therefore, we

have that $U(N(q|p), \varepsilon) = N(q|p)$ across all p and $N(q|p)$, and it follows that $W(N^E, 0, \nu) =$

$W(U^P, 0, \nu)$ and $I(N^E, 0, \nu) = I(U^P, 0, \nu)$. The consequence for the DJA model is that the CHI effect

collapses to zero, and the decomposition (3) can be rewritten as

$$\Delta(0, \nu) = V(0, \nu) - R(0, \nu) = [I(X, 0, \nu) - I(N^E, 0, \nu)] - [I(N, 0, \nu) - I(N^E, 0, \nu)] \qquad (4)$$

It can be shown that $I(X, 0, \nu)$, $I(N, 0, \nu)$, and $I(N^E, 0, \nu)$ are the S-Gini coefficient of pre-fiscal

income, $G(X, \nu)$, the S-Gini coefficient of post-fiscal income, $G(N, \nu)$, and the S-concentration

coefficient of post-fiscal income, $D(N, \nu)$, respectively.[2]

Consequently, $V(0, \nu)$ is equal to the S-Gini Kakwani's (1984) index of vertical effect,

$V^K(\nu) = G(X, \nu) - D(N, \nu)$, and $R(0, \nu)$ is the S-Gini Atkinson (1980), Plotnick (1981), and

Kakwani's (1984) index of reranking, $R^{APK}(\nu) = G(N, \nu) - D(N, \nu)$. The Kakwani's (1984)

decomposition of RE into vertical and horizontal components can be rewritten in S-Gini terms as

$$\Delta(\nu) = V^K(\nu) - R^{APK}(\nu) = [G(X, \nu) - D(N, \nu)] - [G(N, \nu) - D(N, \nu)] \qquad (5)$$

In another special case, where $\nu = 1$, the weights $\omega(p, \nu)$ are equal for all p and the

reranking effect disappears. For $\varepsilon > 0$, the vertical and CHI effect, $V(\varepsilon, 1)$ and $C(\varepsilon, 1)$, become the

indices consistent with the Duclos and Lambert's (2000) model of HI measurement.[3]

3 CALCULATION OF INDICES

3.1. Data Preparation

A typical research uses the following data for a household or family i: (a) unequivalized pre- and post-fiscal incomes, $\dot{X}_i$ and $\dot{N}_i$; (b) survey frequency (or sampling) weights, f_i; and (c) equivalence factor β_i. The equivalized pre- and post-fiscal incomes are $X_i = \dot{X}_i/\beta_i$ and $N_i = \dot{N}_i/\beta_i$ (hereafter we deal only with equivalized incomes, calling them plainly pre- and post-fiscal incomes). The frequency weights are defined as $\phi_i = f_i\beta_i$. Thus, the equivalence factor β_i is employed both for deriving the equivalized income *and* for weighting households of different types.[4]

We form the $3 \times S$ matrix $\mathbf{M}^0$, where S is the number of households in the sample:

$$\mathbf{M}^0 = \begin{bmatrix} X_1 & \dots & X_i & \dots & X_S \\ N_1 & \dots & N_i & \dots & N_S \\ \phi_1 & \dots & \phi_i & \dots & \phi_S \end{bmatrix} \tag{6}$$

To obtain the matrix $\mathbf{M}^x$ ($\mathbf{M}^n$), the columns in $\mathbf{M}^0$ are sorted in increasing order of the values from the first (second) row:

$$\mathbf{M}^x = \begin{bmatrix} X_1^x & \dots & X_i^x & \dots & X_S^x \\ N_1^x & \dots & N_i^x & \dots & N_S^x \\ \phi_1^x & \dots & \phi_i^x & \dots & \phi_S^x \end{bmatrix}; \qquad \mathbf{M}^n = \begin{bmatrix} X_1^n & \dots & X_i^n & \dots & X_S^n \\ N_1^n & \dots & N_i^n & \dots & N_S^n \\ \phi_1^n & \dots & \phi_i^n & \dots & \phi_S^n \end{bmatrix} \tag{7}$$

From $\mathbf{M}^x$ we take out the values X_i^x, N_i^x, and ϕ_i^x, while from $\mathbf{M}^n$ the values N_i^n and ϕ_i^n. and are extracted. Notice that the superscript x (n) denotes that income units are sorted in increasing order of pre-fiscal (post-fiscal) income.

The sample estimates of quantiles p and the weights $\omega(p, v)$ are obtained in the following manner:

$$\tilde{p}_i^x = (2\Sigma)^{-1} \sum_{j=1}^i \left(\phi_j^x + \phi_{j-1}^x \right)$$
$$\tilde{\omega}_i^{x,v} = (\Sigma)^{-1} v(1 - \tilde{p}_i^x)^{v-1} \tag{8}$$

where $\Sigma = \sum_{j=1}^{S} \phi_j^x$ and $\phi_0^x = 0$.

When a large group of pre-fiscal exact equals exists in the sample, one of the inequality indices would be biased if based on the weights $\tilde{\omega}_i^{x,v}$, namely, $\tilde{I}(N_i^x, \varepsilon, v; \tilde{\omega}_i^{x,v})$ from equation (17) (see later discussion). Therefore, we derive a new set of weights, $\hat{\omega}_i^{x,v}$. Assume that income units $i = 1, \dots, Q$ have zero pre-fiscal incomes, i.e., $X_1^x = X_2^x = \cdots = X_Q^x = 0$, and corresponding post-fiscal incomes $N_1^x, N_2^x, \dots, N_Q^x$. The procedure described by equation (8) automatically ascribes to these units the weights that are strictly decreasing in i and $\tilde{p}_i^x$, i.e., $\tilde{\omega}_1^{x,v} > \cdots > \tilde{\omega}_Q^{x,v}$, although all these units have equal pre-fiscal income and rank. Therefore, we replace them with the new set of weights obtained as follows:

$$
\begin{aligned}
\hat{\omega}_i^{x,v} &= \left(\sum_{b=1}^{Q} \phi_b^x\right)^{-1} \sum_{c=1}^{Q} \phi_c^x \cdot \tilde{\omega}_c^{x,v}, && \text{for } i = 1, \dots, Q \\
\hat{\omega}_i^{x,v} &= \tilde{\omega}_i^{x,v}, && \text{for } i > Q
\end{aligned}
\tag{9}
$$

Thus, the original weights $\tilde{\omega}_i^{x,v}$ of pre-fiscal equals are transformed into their group average. An analogous procedure should be applied to other large groups of pre-fiscal equals, if they exist in the sample.

Alternatively, we could use the original weights and randomize the order of income units within each group of exact pre-fiscal equals. This procedure would reduce the bias to an insignificant level, but each possible ordering of income units would still result in different values of $\tilde{I}(N_i^x, \varepsilon, v; \tilde{\omega}_i^{x,v})$. However, for purposes of consistency and transparency, the use of weights $\hat{\omega}_i^{x,v}$ is recommended. For all the other indices derived below, it is irrelevant whether the weights $\tilde{\omega}_i^{x,v}$ or $\hat{\omega}_i^{x,v}$ are employed, because the income vectors they are based upon (namely, X_i^x, $U_{i,\varepsilon}^P$, and N_i^E) have identical values within a group of exact pre-fiscal equals.

Finally, analogously to the above procedures, the estimates $\hat{\omega}_i^{n,v}$ are obtained from ϕ_i^n (for calculation of the inequality index based on N_i^n).

3.2 Indices of Inequality

The following equations show how to obtain utilities, the Gini-Atkinson welfare index, and the inequality index for pre-fiscal incomes X_i^x, when $\varepsilon \neq 1$:

$$U(X_i^x, \varepsilon) = (X_i^x)^{1-\varepsilon}/(1-\varepsilon)$$

$$\widehat{W}(X_i^x, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu}) = \sum_{j=1}^{S} U(X_j^x, \varepsilon) \cdot \phi_j^x \cdot \widehat{\omega}_j^{x,\nu}$$

$$\hat{I}(X_i^x, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu}) = 1 - [(1-\varepsilon)\widehat{W}(X_i^x, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu})]^{\frac{1}{1-\varepsilon}}/\hat{\mu}(X_i^x) \tag{10}$$

where $\hat{\mu}(X_i^x) = (\Sigma)^{-1} \sum_{j=1}^{S} \phi_j^x \cdot X_j^x$ is the mean pre-fiscal income. To shorten the presentation, the formulas referring to the case where $\varepsilon = 1$ are omitted. Analogously, the utilities and indices for post-fiscal incomes N_i^n are obtained, as shown by equation (16) in the Appendix.

$\hat{I}(X_i^x, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu})$ and $\hat{I}(N_i^n, \varepsilon, \nu; \widehat{\omega}_i^{n,\nu})$ from (10) and (16) are the sample estimates of the indices of pre- and post-fiscal income inequality, $I(X)$ and $I(N)$. As equation (3) indicates, the application of the DJA model requires the estimates of two other indices, $I(N^E)$ and $I(U^P)$, derived from EPIs, $N^E(p)$, and EPUs, $U^P(p, \varepsilon)$.

To obtain the sample estimates of EPUs, $U_{i,\varepsilon}^P$, we should smooth a dataset $\{X_i^x, U(N_i^x, \varepsilon); \phi_i^x\}_{i=1}^{S}$. For each value of ε, we estimate the regression relationship $U(N_i^x, \varepsilon) = m^{P,\varepsilon}(X_i^x) + \sigma_i$, to obtain the approximation $\tilde{m}^{P,\varepsilon}(\cdot)$. Subsequently, the vector of fitted values is calculated as $U_{i,\varepsilon}^P = \tilde{m}^{P,\varepsilon}(X_i^x)$.

However, the following identity says that the whole procedure of estimating EPUs can be circumvented, saving the practitioner's time and energy in sensitivity analysis using multiple scenarios for ν and ε. Genuinely, the sample estimate of $I(U^P)$ is equal to $\hat{I}(U_{i,\varepsilon}^P, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu})$ from equation (20), but it can be derived more simply by the inequality index $\hat{I}(N_i^x, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu})$ from equation (18), because of the following equality:

$$\hat{I}(U_{i,\varepsilon}^P, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu}) = \hat{I}(N_i^x, \varepsilon, \nu; \widehat{\omega}_i^{x,\nu}) \tag{11}$$

To understand why (11) holds, recall that $U^P(p, \varepsilon) = \int_0^1 U(N(q|p), \varepsilon)dq$ and notice that the theoretical values $N(q|p)$ are represented by the sample values N_1^x. Imagine the population consisting of two groups of exact pre-fiscal equals: B income units have pre-fiscal income $X_{(1)}^x$ and post-fiscal incomes $N_1^x, \dots, N_B^x$, while F income units have pre-fiscal income $X_{(2)}^x > X_{(1)}^x$ and post-fiscal incomes $N_{B+1}^x, \dots, N_{B+F}^x$. Their rank-dependent weights are $\widehat{\omega}_1^{x,v} = \cdots = \widehat{\omega}_B^{x,v} = \widehat{\omega}_{(1)}^{x,v}$ and $\widehat{\omega}_{B+1}^{x,v} = \cdots = \widehat{\omega}_{B+F}^{x,v} = \widehat{\omega}_{(2)}^{x,v}$. For given ε, the values of expected post-fiscal utilities, $U_{i,\varepsilon}^P$, are obtained as averages of utilities $U_{(1),\varepsilon}^P = B^{-1} \sum_{j=1}^B U(N_j^x, \varepsilon)$ and $U_{(2),\varepsilon}^P = F^{-1} \sum_{j=B+1}^{B+F} U(N_j^x, \varepsilon)$. The welfare is then obtained by (20) as

$$\widehat{W}(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v}) = B \cdot \widehat{\omega}_{(1)}^{x,v} \cdot B^{-1} \sum_{j=1}^B U(N_j^x, \varepsilon) + F \cdot \widehat{\omega}_{(2)}^{x,v} \cdot F^{-1} \sum_{j=B+1}^{B+F} U(N_j^x, \varepsilon)$$

$$= \widehat{\omega}_{(1)}^{x,v} \cdot \sum_{j=1}^B U(N_j^x, \varepsilon) + \widehat{\omega}_{(2)}^{x,v} \cdot \sum_{j=B+1}^{B+F} U(N_j^x, \varepsilon)$$

Observe that according to (18) we would obtain the identical result for $\widehat{W}(N_i^n, \varepsilon, v; \widehat{\omega}_i^{n,v})$. Consequently, $\hat{I}(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v})$ and $\hat{I}(N_i^x, \varepsilon, v; \widehat{\omega}_i^{x,v})$ are also identical.

3.3 Estimation of Expected Post-fiscal Incomes and Utilities

Unlike the estimation of EPUs, the evaluation of EPIs cannot be avoided. To obtain the sample estimates of $N^E(p)$, we must smooth a dataset $\{X_i^x, N_i^x; \phi_i^x\}_{i=1}^S$, i.e., approximate the mean response curve m^E in the regression relationship $N_i^x = m^E(X_i^x) + \delta_i$. The estimates of EPIs are then obtained as $N_i^E = \tilde{m}^E(X_i^x)$, where $\tilde{m}^E(X_i^x)$ is the approximation of m^E.

The estimation of EPIs represents the greatest challenge in the implementation of the DJA model. Although parametric models (such as polynomial regression) can be appropriate for some datasets, it is better to rely on non-parametric approaches, assuming no a priori functional relationship between post- and pre-fiscal incomes. One such approach is the 'kernel-weighted local polynomial regression' (KWLPR). A description of the method can be found in Fan and Gijbels

(1996), Wand and Jones (1995), Keele (2008), and Härdle (1990), while the software applications include Stata 12 (function *lcpoly*), R (function *loess*, package *lokern*, etc.), and XploRe (function *lpregxest*).

The choice of the degree of polynomial (p), the type of the kernel function, and the size of the kernel half-bandwidth rests on the analyst. For $p = 0$, KWLPR becomes the 'Nadaraya-Watson estimator' (NWE), while for $p = 1$ we obtain the 'local linear estimator' (LLE). Fan and Gijbels (1996) explain that the odd degree polynomials achieve the best balance between bias and variability and automatically correct the boundary problem.

Another interesting smoothing technique came to light during the research: the 'Fourier series in trigonometric form' (FSTF), which is a sum of sine and cosine functions describing a periodic signal (Faunt and Johnson 1992). The estimation procedure is programmed in Matlab R2011b's Curve Fitting Toolbox 3.2, which contains several other fitting methods, such as smoothing splines.

DAD-DJA and supporting documentation[5] do not inform us which fitting method is used to obtain EPIs for estimation of DJA indices. However, DAD incorporates a separate module, 'Non Parametric Regression', enabling us to estimate EPIs independently of DAD-DJA. Two basic methods are offered: NWE and LLE (henceforth DAD-NWE and DAD-LLE). Experimentation with different options and choices offered by the module revealed that in estimating EPIs DAD-DJA in fact employs DAD-LLE, using the default set of parameters.

Before moving further, we offer the following advice to help judge whether the estimates N_i^E are appropriate for use in the DJA model implementation.

(a)   Although the fitting methods and their software implementations ensure optimality in the statistical sense, the analyst still has the freedom and the responsibility to change some of the parameters or the whole estimation method if the results contradict her/his knowledge of the appropriate shape of the EPIs curve. An example is a too 'wiggly' curve, in which case we have to

'stretch' it, perhaps by raising the kernel half-bandwidth. Another example may be the existence of certain kinks or local minimums (maximums) we are aware of, which are not reflected by the EPIs estimate.

(b)   For certain data points the programmed fitting procedures may produce irregular results. Some software tools are 'smart' in such cases, leaving a blank space instead of the estimate, while others are not. Anyway, if this happens we should fill in the corresponding estimate manually, using the best-guess approach.

(c)   A simple preliminary test of the correctness of the approximation $\tilde{m}^E(X_i^x)$ is to check whether the mean value of the estimated values N_i^E is approximately equal to the mean of the sample values N_i^x, i.e., if

$$\hat{\mu}\left(N_i^E\right) = \hat{\mu}(N_i^x) \tag{12}$$

(d)   The discussion in section 2 indicated that when $\varepsilon = 0$, we have that $U(N_i^x, 0) = N_i^x$, and $U_{i,\varepsilon}^P = N_i^E$. Therefore, equation (11) becomes

$$\hat{I}\left(N_i^E, 0, v; \widehat{\omega}_i^{x,v}\right) = \hat{I}\left(U_{i,\varepsilon}^P, 0, v; \widehat{\omega}_i^{x,v}\right) = \hat{I}\left(N_i^x, 0, v; \widehat{\omega}_i^{x,v}\right) \tag{13}$$

From equation (13) follows another test: the inequality indices $\hat{I}\left(N_i^E, 0, v; \widehat{\omega}_i^{x,v}\right)$, obtained by equation (19), for different values of parameter v should be (approximately) equal to the inequality indices obtained for $\hat{I}\left(N_i^x, 0, v; \widehat{\omega}_i^{x,v}\right)$. Otherwise, the estimates of the DJA indices would be biased, and we should try to obtain an alternative configuration of $\tilde{m}^E(X_i^x)$.

3.4 Decompositions

Having defined all the indices needed, we can present RE and its decompositions in terms of sample estimate formulas. RE is obtained as $\hat{\Delta} = \hat{I}(X_i^x) - \hat{I}(N_i^n)$. According to the DJA model from (3), RE is decomposed as follows:

$$\hat{\Delta} = \hat{V} - \hat{C} - \hat{R} =$$

$$= \left[\hat{I}(X_i^x) - \hat{I}(N_i^E)\right] - \left[\hat{I}(U_{i,\varepsilon}^P) - \hat{I}(N_i^E)\right] - \left[\hat{I}(N_i^n) - \hat{I}(U_{i,\varepsilon}^P)\right] =$$
$$= \left[\hat{I}(X_i^x) - \hat{I}(N_i^E)\right] - \left[\hat{I}(N_i^x) - \hat{I}(N_i^E)\right] - \left[\hat{I}(N_i^n) - \hat{I}(N_i^x)\right] \tag{14}$$

where the last row in equation (14) arrives from the property (11), by which $\hat{I}(U_{i,\varepsilon}^P) = \hat{I}(N_i^x)$. The

differences in the brackets, i.e., $\hat{V} = \hat{I}(X_i^x) - \hat{I}(N_i^E)$, $\hat{C} = \hat{I}(U_{i,\varepsilon}^P) - \hat{I}(N_i^E) = \hat{I}(N_i^x) - \hat{I}(N_i^E)$, and

$\hat{R} = \hat{I}(N_i^n) - \hat{I}(U_{i,\varepsilon}^P) = \hat{I}(N_i^n) - \hat{I}(N_i^x)$, are respectively the sample estimates of the vertical, CHI, and

reranking effects of the DJA model. Note that, if $\tilde{I}(N_i^x)$ is used instead of $\hat{I}(N_i^x)$, the CHI and

reranking effects would be biased. Furthermore, if the estimates N_i^E are inappropriate, the vertical

and CHI effect would be biased.

Setting $\varepsilon = 0$ and following (4) and (5), we can calculate the S-Gini K84 decomposition as

$$\hat{\Delta} = \widehat{V^K}(\nu) - \widehat{R^{APK}}(\nu) =$$
$$\left[\hat{I}(X_i^x, 0, \nu; \widehat{\omega}_i^{x,\nu}) - \hat{I}(N_i^x, 0, \nu; \widehat{\omega}_i^{x,\nu})\right] - \left[\hat{I}(N_i^n, 0, \nu; \widehat{\omega}_i^{n,\nu}) - \hat{I}(N_i^x, 0, \nu; \widehat{\omega}_i^{x,\nu})\right] \tag{15}$$

3.5 Simple Hypothetical Example

We return to the example of four hypothetical households from section 2 to illustrate how the DJA

model implementation procedures work. There are two groups of pre-fiscal equals in the sample: A

and B with pre-fiscal income of 10 each belong to the lower quantile, whereas C and D with pre-

fiscal income of 20 each belong to the upper quantile of pre-fiscal income distribution.

The first column in Table 1 shows the 'original' weights $\widetilde{\omega}_i^{x,\nu}$ obtained by (8) for $\nu = 2$;

observe that A (C) obtains larger weight than B (D), although they belong to same pre-fiscal

quantile. Therefore, analogously to the procedure from equation (9), we obtain the new set of

weights, $\widehat{\omega}_i^{x,\nu}$: for A and B (C and D) we have $\widehat{\omega}_1^x = \widehat{\omega}_2^x = (\widetilde{\omega}_1^x + \widetilde{\omega}_2^x)/2$ $[\widehat{\omega}_3^x = \widehat{\omega}_4^x = (\widetilde{\omega}_3^x + \widetilde{\omega}_4^x)/2]$.

Table 1

As equation (11) explains, the estimate of $I(U^P)$ can be obtained in two ways: by $\hat{I}(U_{i,\varepsilon}^P)$ or

by $\hat{I}(N_i^x)$. In the former case, we first obtain the values $U_{i,\varepsilon}^P$, which are the sample estimates of

$U^P(p, \varepsilon) = \int_0^1 U(N(q|p), \varepsilon)dq$. The values $N(q|p)$ for the lower (upper) quantile are represented by

N_1^x and N_2^x (N_3^x and N_4^x), and the corresponding utilities are $U(N_1^x)$ and $U(N_2^x)$ [$U(N_3^x)$ and $U(N_4^x)$]. The expected post-fiscal utility at each quantile is simply the average utility of units belonging to the corresponding quantile: $U_1^P = U_2^P = [U(N_1^x) + U(N_2^x)]/2$ and $U_3^P = U_4^P = [U(N_3^x) + U(N_4^x)]/2$, for the lower and the upper quantiles, respectively.

According to (20), for $v = 2$ and $\varepsilon = 0.5$ we have that

$$\widehat{W}\left(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v}\right) = 2 \cdot U_1^P \cdot \widehat{\omega}_1^x + 2 \cdot U_3^P \cdot \widehat{\omega}_3^x = 7.21$$

and $\hat{I}\left(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v}\right) = 0.133$. Substituting back previously obtained utility terms into the expression for welfare, we obtain

$$\widehat{W}\left(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v}\right) = [U(N_1^x) + U(N_2^x)] \cdot \widehat{\omega}_1^x + [U(N_3^x) + U(N_4^x)] \cdot \widehat{\omega}_3^x = 7.21$$

which is identical to the result that would be obtained by (18):

$$\widehat{W}\left(N_i^x, \varepsilon, v; \widehat{\omega}_i^{x,v}\right) = U(N_1^x) \cdot \widehat{\omega}_1^x + U(N_2^x) \cdot \widehat{\omega}_2^x + U(N_3^x) \cdot \widehat{\omega}_3^x + U(N_4^x) \cdot \widehat{\omega}_4^x = 7.21,$$

Thus, our hypothetical example confirms the identity (11). On the other hand, indices based on the 'wrong' weights, $\widetilde{\omega}_i^{x,v}$, would produce quite a different picture. By (17) we have

$$\widetilde{W}\left(N_i^x, \varepsilon, v; \widetilde{\omega}_i^{x,v}\right) = U(N_1^x) \cdot \widetilde{\omega}_1^x + U(N_2^x) \cdot \widetilde{\omega}_2^x + U(N_3^x) \cdot \widetilde{\omega}_3^x + U(N_4^x) \cdot \widetilde{\omega}_4^x = 6.89,$$

and $\tilde{I}\left(N_i^x, \varepsilon, v; \widetilde{\omega}_i^{x,v}\right) = 0.210$.

The estimate of $I(N^E)$ is obtained as $\hat{I}\left(N_i^E, \varepsilon, v; \widehat{\omega}_i^{x,v}\right)$ from (19), which is in turn based on sample estimates N_i^E of $N^E(p) = \int_0^1 N(q|p)dq$. In this example we can simply calculate $N_1^E = N_2^E = (N_1^x + N_2^x)/2$ and $N_3^E = N_4^E = (N_3^x + N_4^x)/2$. According to (19),

$$\widehat{W}\left(N_i^E, \varepsilon, v; \widehat{\omega}_i^{x,v}\right) = 2 \cdot N_1^E \cdot \widehat{\omega}_1^x + 2 \cdot N_3^E \cdot \widehat{\omega}_3^x = 7.32,$$

and $\hat{I}\left(N_i^E, \varepsilon, v; \widehat{\omega}_i^{x,v}\right) = 0.108$.

All inequality indices for $v = 2$ and $\varepsilon = 0.5$ are presented in the first column of Table 2, together with the DJA decomposition results. Although the vertical effect of the hypothetical system is positive ($\hat{V} = 0.081$), the redistributive effect is negative ($\hat{\Delta} = -0.052$), because the CHI effect ($\hat{C} = 0.026$) and especially the reranking effect ($\hat{R} = 0.107$) are very large.

Table 2

Another set of CHI and reranking effects is derived using $\tilde{I}(N_i^x)$, which is based on the weights $\tilde{\omega}_i^{x,v}$. They show a completely different picture of the relative contributions of CHI ($\tilde{C} = 0.102$ vs. $\hat{C} = 0.026$) and reranking ($\tilde{R} = 0.030$ vs. $\hat{R} = 0.107$) to the overall HI. We have indicated that the weights $\hat{\omega}_i^{x,v}$ are the 'right ones', but to demonstrate this in our example, we have to obtain the indices for $v = 2$ and $\varepsilon = 0$, shown in the second column of Table 2.

Recall that equation (13) says that the inequality indices based on N_i^x, $U_{i,\varepsilon}^P$, and N_i^E must be equal when $\varepsilon = 0$; this is true for $\hat{I}(N_i^x) = 0.100$, but not for $\tilde{I}(N_i^x) = 0.183$. The difference $\tilde{I}(N_i^x) - \hat{I}(N_i^x) = 0.083$ presents by how much the CHI effect, which should be zero when $\varepsilon = 0$, is overestimated if the weights $\tilde{\omega}_i^{x,v}$ are used in computation of the inequality index for post-fiscal incomes N_i^x.

Finally, we look at how DAD-DJA deals with this small hypothetical case. The DAD supporting documentation tells us that the estimate of $I(U^P)$ is obtained by the index based on N_i^x; its value for $v = 2$ and $\varepsilon = 0$ (0.183) is identical to $\tilde{I}(N_i^x)$. This indicates that DAD-DJA does not envisage the presence of exact equals in the sample. The third column in Table 2 shows the other results. The index $\hat{I}(N_i^E)$ diverges highly from our estimate (0.500 vs. 0.100), but this may be due to the small sample size. We deal with the DAD-DJA estimates of EPIs in the next section.

4 APPLICATION: CROATIAN TAX-BENEFIT SYSTEM

4.1 Data

We analyze the fiscal system consisting of social security contributions (SSC) for the pension, health, and unemployment insurance funds, personal income tax and surtax (PITS), public pensions, and cash social benefits.[6] The data on incomes come from the Croatian household budget survey (Anketa o potrošnji kućanstava; APK) for 2008, whose sample contains 3,108 households. Since APK registers only net incomes of household members, the amounts of pre-fiscal income, PITS, and SSC are obtained by a microsimulation model.

Post-fiscal income of a household i is obtained as $\dot{N}_i = \dot{X}_i - \dot{T}_i + \dot{B}_i$, where $\dot{X}_i$, $\dot{T}_i$, and $\dot{B}_i$ are pre-fiscal income, the sum of all taxes paid, and the sum of all benefits received. To obtain X_i and N_i, $\dot{X}_i$ and $\dot{N}_i$ are deflated by the equivalence factor β_i according to the 'modified OECD scale', $\beta_i = 1 + 0.5(a_i - 1) + 0.3c_i$, where a_i and c_i are numbers of adults and children in household i.

Before analyzing the results of the DJA decomposition, we observe the features of the data set. The dots in the scattergram (Figure 1) are the post-fiscal and pre-fiscal incomes of sample income units, expressed in terms of the mean pre-fiscal income (*mpfi*). The full line shows EPIs obtained by KWLPR (see next section for details on estimation). The dotted line represents the cumulative density, which tells us, for each pre-fiscal income *X*, the proportion of all income units having pre-fiscal income below *X* (on the right axis). We can observe that quite a large proportion of units, about 7 percent, have zero pre-fiscal income (group A), while the next 13 percent of units have pre-fiscal income below 10 percent of *mpfi* (group B).

Figure 1

The mean post-fiscal incomes of groups A and B are 64 and 54 percent of *mpfi*, respectively. Observe that the EPIs curve is decreasing on the interval [0, 0.1]. The following three facts taken together can explain the curious feature that the mean post-fiscal income is decreasing. First, for the

majority of pensioners' households a public pension is the only source of income; since public pensions are benefits in the current scenario, the pre-fiscal income of most pensioners' households is zero. Second, majority of households with zero pre-fiscal income (group A) are pensioners' households. Third, pensions are on average higher than other social benefits.

4.2 Estimation of Expected Post-fiscal Incomes and the Decomposition

The indices of the DJA decomposition are estimated by three models, using three different fitting methods described in section 3.3.

In model A, EPIs are estimated by KWLPR programmed in Stata 12. Following Bilger (2008), we use the 3rd degree local polynomials, employing the Epanechnikov kernel. The optimal half-bandwidth of the kernel obtained by the program was equal to 6.7 percent of *mpfi* and it was increased by one half. In model B, EPIs are obtained using FSTF programmed in Matlab R2011b's Curve Fitting Toolbox 3.2. The number of harmonics is set to 7; the "Trust-Region" algorithm is employed with the robust fitting option turned off. In both models the top five pre-fiscal income units are excluded from the fitting process, and their values of N_i^E are set to the values of N_i^x. The size of the half-bandwidth in model A and the number of harmonics in model B are chosen to minimize the bias $\hat{I}(N_i^x, 0, v; \widehat{\omega}_i^{x,v}) - \hat{I}(N_i^E, 0, v; \widehat{\omega}_i^{x,v})$. For the estimates of $I(U^P)$, we used $\hat{I}(N_i^x, \varepsilon, v; \widehat{\omega}_i^{x,v})$ with weights $\widehat{\omega}_i^{x,v}$ obtained by equation (9).

The aim of model C is to replicate the results obtained by DAD-DJA. We employ DAD-LLE to estimate EPIs, with all observations included in the fitting process. To estimate $I(U^P)$, DAD-DJA also uses the index based on N_i^x, but does not envisage the possibility of pre-fiscal exact equals. To play down the bias in the calculation of reranking and CHI effects, we randomize the order of income units within the group of zero pre-fiscal equals; these data are then put into DAD-DJA to obtain the original estimates, while in the replication we calculate $\tilde{I}(N_i^x, \varepsilon, v; \widetilde{\omega}_i^{x,v})$.

Before moving on to the results, let us look at the shapes of the different EPIs curves, shown in Figure 2, concentrating first on the bottom part of the income distribution. While A and B both reflect the initial fall in expected post-fiscal income, discussed above, C does not, i.e., its EPIs curve is rather flat on the whole interval. For pre-fiscal income of zero all estimates are roughly the same, but in the pre-fiscal income interval [0.025, 0.42] of *mpfi*, C's EPIs lie above those estimated by A and B. On the pre-fiscal income interval [0, 0.5] of *mpfi* the mean of EPIs obtained by A (B) is 0.5764 (0.5775) of *mpfi*, which is very close to the mean post-fiscal income for actual values, equal to 0.5762. On the other hand, the mean of EPIs obtained by C is 0.5881, or 2 percent above the actual mean. This suggests that C overestimates EPIs for the lowest incomes. For pre-fiscal incomes above 0.5 of *mpfi*, the EPIs of B and C are almost identical, while the EPIs curve of A is "more flexible" and intertwining the other two curves.

Figure 2

Models A and B convincingly pass the test from equation (12), as the ratios $\hat{\mu}(N_i^E)/\hat{\mu}(N_i^x)$ in Table 3 are 0.999732 and 1.0, respectively. On the other hand, for method C, $\hat{\mu}(N_i^E)$ is 1.75 percent higher than $\hat{\mu}(N_i^x)$. This is partly the consequence of the earlier noticed overestimation on the interval [0, 0.5] of *mpfi*. However, there is another feature that is particularly odd: the estimate N_i^E for the two income units with top pre-fiscal incomes are 2.4 and 6.4 times larger than their respective actual post-fiscal incomes N_i^x! In this case we can talk about a 'bug' in DJA-LLE, which seriously damages the estimate of $\hat{\mu}(N_i^E)$, which will lead to biased estimates of vertical and CHI effects, as we will soon observe. Table 3 shows the results of the DJA decomposition for $\nu = 2$ and $\varepsilon = 0$. All three models obtain equal values of $\hat{I}(X_i^x)$ and $\hat{I}(N_i^n)$. Models A and B obtain the value of $\hat{I}(N_i^x)$ equal to 0.244337, which is almost insignificantly different from the value of $\tilde{I}(N_i^x)$ obtained by model C, thanks to randomizing the order of income units within the group of zero pre-fiscal equals. If the units within the group of zero pre-fiscal equals were, by chance, sorted in increasing

(or decreasing) order of post-fiscal income, the difference $\left|\tilde{I}(N_i^x) - \hat{I}(N_i^x)\right|$ for the given data set could be as high as 0.001183. The reranking and CHI effects in model C could be seriously biased.

Table 3

The estimates $\hat{I}\left(N_i^E\right)$ obtained by A and B are close to the value of $\hat{I}(N_i^x)$, as expected from equation (13); the differences $\hat{I}(N_i^x) - \hat{I}\left(N_i^E\right)$ are -0.000203 and -0.000180, or between -0.09 and 0.08 percent of the corresponding RE $(\hat{\Delta})$. For the model C, the difference $\hat{I}(N_i^x) - \hat{I}\left(N_i^E\right)$ is no less than -0.007104, or -3.3 percent of RE, meaning that the bias produced by C is about 35 times larger than the bias of A and B.

Table 4

Model C thus underestimates the vertical effect by about 3 percent of RE for $\nu = 2$ and $\varepsilon = 0$. This underestimation is even larger (somewhat smaller) for $\nu = 1.5$ and $\varepsilon = 0$ ($\nu = 3$ and $\varepsilon = 0$) and amounts to 5.9 (2.4) percent of RE, as Table 4 indicates. Among the scenarios shown in Table 4, the differences in the estimates of vertical effect obtained by models C and A (B) are lowest when $\nu = 2$ and $\varepsilon = 0.9$, equaling 0.8 (0.9) percent of RE.

5 CONCLUSION

Models decomposing the redistributive effect of fiscal systems into vertical and horizontal effects are extensively used by practitioners. The Duclos, Jalbert, and Araar (2003) model, despite its advantages over some other models, such as the Kakwani's (1984) and the Aronson, Johnson and Lambert's (1994) decompositions of RE, has not yet been broadly employed in empirical research. The reason may be the relatively complex implementation procedure, which involves non-parametric methods in estimation of expected post-fiscal incomes.

To override these estimation and calculation difficulties, the designers of the software DAD have incorporated a module for estimation of the DJA model indices, here referred to as DAD-DJA.

However, as the application data on the Croatian tax-benefit system indicates, DAD-DJA produces somewhat inaccurate estimates of EPIs, resulting in biased values of DJA model indices. This paper carefully explains the estimation procedures needed to obtain the indices of the DJA model, and the problems occurring in DAD-DJA implementation.

The estimates of expected post-fiscal incomes are obtained by two fitting methods, kernel-weighted local polynomial regression and Fourier series in trigonometric form. Both achieve reasonable fit of the data at stake, unlike the method built into DAD-DJA, which seems to overestimate EPIs at the bottom region of pre-fiscal income distribution. Furthermore, we have realized that the fitting procedure in DAD-DJA contains a 'bug', producing unreasonably high estimates of EPIs for the top pre-fiscal income units in the sample.

We have shown how the estimation of EPUs can be circumvented, saving a practitioner time when doing multiple-scenario analysis. Instead of estimating EPUs for each different value of parameter ε, the index of inequality based on EPUs can be obtained simply by using post-fiscal incomes ordered according to pre-fiscal incomes. In this procedure, however, caution must be taken in the presence of large groups of exact pre-fiscal equals: they should all be ascribed identical ranking weights. Otherwise, an estimate of reranking effect would be biased.

REFERENCES

Araar, Abdelkrim, and Jean-Yves Duclos. 2003. An Atkinson-Gini Family of Social Evaluation Functions. *Economics Bulletin* 3 (19): 1–16.

Araar, Abdelkrim, and Jean-Yves Duclos. 2006. An Atkinson-Gini Family of Social Evaluation Functions: Theory and Illustration Using Data from the Luxembourg Income Study. *Public Finance* 54 (3): 191–209.

Aronson, J. Richard, Paul Johnson, and Peter J. Lambert. 1994. Redistributive effect and unequal income tax treatment. *The Economic Journal* 104: 262–270.

Atkinson, Anthony B. 1970. On the Measurement of Inequality. *Journal of Economic Theory* 2: 244–263.

Atkinson, Anthony B. 1980. Horizontal Equity and the Distribution of the Tax Burden. In *The Economics of Taxation*, Henry J. Aaron, and Michael J. Boskins, eds., 3–18. Washington D.C.: Brookings Institution.

Bilger, Marcel. 2008. Progressivity, horizontal inequality and reranking caused by health system financing: A decomposition analysis for Switzerland. *Journal of Health Economics* 27: 1582–1593.

Dardanoni, Valentino, and Peter J. Lambert. 2001. Horizontal Inequity Comparisons. *Social Choice and Welfare* 18: 799–816.

Donaldson, David, and John A. Weymark. 1980. A Single-Parameter Generalization of Gini indices of Inequality. *Journal of Economic Theory* 22: 67–86.

Duclos, Jean-Yves, Abdelkrim Araar, and Carl Fortin. 2010. DAD: A software for Distributive Analysis / Analyse Distributive. MIMAP programme, International Development Research Centre, Government of Canada, and CIRPÉE, Université Laval (version: DAD 4.6 of October 2010).

Duclos, Jean-Yves, Vincent Jalbert, and Abdelkrim Araar. 2003. Classical horizontal inequity and reranking: an integrated approach. *Research on Economic Inequality* 10: 65–100.

Duclos, Jean-Yves, and Peter J. Lambert. 2000. A Normative and Statistical Approach to Measuring Classical Horizontal Inequity. *Canadian Journal of Economics* 33: 87–113.

Ebert, Udo, 1997. Social Welfare When Needs Differ: An Axiomatic Approach. Economica 64: 233–44.

Ebert, Udo. 1999. Using equivalent income of equivalent adults to rank income distributions. *Social Choice and Welfare* 16: 233–258.

Fan, Jianqing, and Gijbels, Irène. 1996. Local polynomial modelling and its applications. London: Chapman and Hall.

Faunt, Lindsay M., and Michael L. Johnson. 1992. Analysis of discrete, time-sampled data using Fourier series method. *Methods in Enzymology* 210: 340–56.

Härdle, Wolfgang. 1990. Applied Nonparametric Regression. Cambridge University Press.

Kakwani, Nanak C. 1984. On the measurement of tax progressivity and redistributive effect of taxes with applications to horizontal and vertical equity. *Advances in Econometrics* 3: 149–168.

Keele, Luke. 2008. Semiparametric Regression for the Social Sciences. London: Wiley & Sons.

Plotnick, Robert. 1981. A Measure of Horizontal Equity. *The Review of Economics and Statistics* 63: 283–288.

Wand, Matt P. and Chris M. Jones. 1995. Kernel Smoothing. London: Chapman and Hall.

Yitzhaki, Shlomo. 1983. On an Extension of the Gini Inequality Index. *International Economic Review* 24: 617–628.

Yitzhaki, Shlomo, and Ingram Olkin. 1991. Concentration curves and concentration indices. *Lecture Notes - Monograph Series* 19: 380–392.

APPENDIX 1 SAMPLE ESTIMATES OF INEQUALITY INDICES

The sample estimates of Atkinson-Gini Inequality indices based on N_i^n, N_i^x, N_i^E and U_i^P are obtained in the following equations:

$$U(N_i^n, \varepsilon) = (N_i^n)^{1-\varepsilon}/(1-\varepsilon)$$
$$\widehat{W}(N_i^n, \varepsilon, v; \widehat{\omega}_i^{n,v}) = \sum_{j=1}^{S} U(N_j^n, \varepsilon) \cdot \phi_j^n \cdot \widehat{\omega}_j^{n,v}$$
$$\widehat{I}(N_i^n, \varepsilon, v; \widehat{\omega}_i^{n,v}) = 1 - [(1-\varepsilon)\widehat{W}(N_i^n, \varepsilon, v; \widehat{\omega}_i^{n,v})]^{\frac{1}{1-\varepsilon}}/\hat{\mu}(N_i^n) \tag{16}$$

$$U(N_i^x, \varepsilon) = (N_i^x)^{1-\varepsilon}/(1-\varepsilon)$$
$$\widetilde{W}(N_i^x, \varepsilon, v; \widetilde{\omega}_i^{x,v}) = \sum_{j=1}^{S} U(N_j^x, \varepsilon) \cdot \phi_j^x \cdot \widetilde{\omega}_j^{x,v}$$
$$\widetilde{I}(N_i^x, \varepsilon, v; \widetilde{\omega}_i^{x,v}) = 1 - [(1-\varepsilon)\widetilde{W}(N_i^x, \varepsilon, v; \widetilde{\omega}_i^{x,v})]^{\frac{1}{1-\varepsilon}}/\hat{\mu}(N_i^x) \tag{17}$$

$$\widehat{W}(N_i^x, \varepsilon, v; \widehat{\omega}_i^{x,v}) = \sum_{j=1}^{S} U(N_j^x, \varepsilon) \cdot \phi_j^x \cdot \widehat{\omega}_j^{x,v}$$
$$\widehat{I}(N_i^x, \varepsilon, v; \widehat{\omega}_i^{x,v}) = 1 - [(1-\varepsilon)\widehat{W}(N_i^x, \varepsilon, v; \widehat{\omega}_i^{x,v})]^{\frac{1}{1-\varepsilon}}/\hat{\mu}(N_i^x) \tag{18}$$

$$U(N_i^E, \varepsilon) = (N_i^E)^{1-\varepsilon}/(1-\varepsilon)$$
$$\widehat{W}(N_i^E, \varepsilon, v; \widehat{\omega}_i^{x,v}) = \sum_{j=1}^{S} U(N_j^E, \varepsilon) \cdot \phi_j^x \cdot \widehat{\omega}_j^{x,v}$$
$$\widehat{I}(N_i^E, \varepsilon, v; \widehat{\omega}_i^{x,v}) = 1 - [(1-\varepsilon)\widehat{W}(N_i^E, \varepsilon, v; \widehat{\omega}_i^{x,v})]^{\frac{1}{1-\varepsilon}}/\hat{\mu}(N_i^E) \tag{19}$$

$$\widehat{W}(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v}) = \sum_{j=1}^{S} U_{i,\varepsilon}^P \cdot \phi_j^x \cdot \widehat{\omega}_j^{x,v}$$
$$\widehat{I}(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v}) = 1 - [(1-\varepsilon)\widehat{W}(U_{i,\varepsilon}^P, \varepsilon, v; \widehat{\omega}_i^{x,v})]^{\frac{1}{1-\varepsilon}}/\hat{\mu}(N_i^x) \tag{20}$$

where $\hat{\mu}(N_i^n)$, $\hat{\mu}(N_i^x)$, and $\hat{\mu}(N_i^E)$ are means of post-fiscal income variables, equal to $\hat{\mu}(N_i^n) = (\varSigma)^{-1} \sum_{j=1}^{S} \phi_j^n \cdot N_j^n$, $\hat{\mu}(N_i^x) = (\varSigma)^{-1} \sum_{j=1}^{S} \phi_j^x \cdot N_j^x$, and $\hat{\mu}(N_i^E) = (\varSigma)^{-1} \sum_{j=1}^{S} \phi_j^x \cdot N_j^E$, respectively. It is clear that $\hat{\mu}(N_i^x) = \hat{\mu}(N_i^n)$, because N_i^x and N_i^n contain the same sample values, only differently sorted.

ENDNOTES

[1] Araar and Duclos (2003, 2006) describe the properties of AGF based inequality indices: "Income inequality aversion is captured by decreasing marginal utilities, and aversion to rank inequality is captured by rank-dependent ethical weights, thus providing an ethically-flexible dual basis for the assessment of inequality and equity" (Araar and Duclos 2006, 192). Furthermore, it is shown that AGF is the only family of social evaluation functions "to obey a set of popular axioms in the income distribution literature" (Araar and Duclos 2006, 204).

[2] Independent proof of this relationship can be found in Yitzhaki and Olkin (1991), who derive the "relative concentration curve" of post-fiscal income N with respect to pre-fiscal income X as $C(N, x, p) = (\mu^N)^{-1} \int_{-\infty}^{F_X^{-1}(p)} m(t) dF_X(t)$, where $m(x) = E\{N|X = x\}$ corresponds to $N^E(p)$. Duclos and Araar (2006) present the same concentration curve as $C(N, x, p) = (\mu^N)^{-1} \int_0^p N^E(r) dr$, from which the S-Gini concentration coefficient is obtained as $D(N, v) = \int_0^1 \left[\left(p - C(N, x, p) \right) \right] \varpi(p, v)$, where $\varpi(p, v) = v(v-1)(1-p)^{v-2}$ are rank-dependent weights.

[3] These authors have derived their indices using the "cost of inequality" approach, compared with the "change of inequality" approach used in this paper. Duclos, Jalbert, and Araar (2003) employ both approaches.

[4] According to Ebert (1997, 1999) this is the right approach to investigate the concepts of Lorenz dominance, social welfare function, and progressive transfers when populations are heterogeneous. Using the number of 'real' household members, u_i, instead of the number of 'equivalent' members, β_i, leads to "some unpleasant and unsatisfactory paradoxa or impossibility results". The usual objection to this approach is that "not all persons have the same weight and significance", which contradicts the democratic principles; for the rebuttal of this objection see Ebert (1999, 251).

[5] Available online at: http://132.203.59.36/DAD/manual/help_44.htm.

Table 1

Hypothetical population: weight, incomes, and utilities

#	$\widetilde{\omega}_i^x$	$\widehat{\omega}_i^x$	X_i^x	N_i^x	N_i^n	N_i^E	$U(X_i^x)$	$U(N_i^x)$	$U(N_i^n)$	$U(N_i^E)$	$U_{i,\varepsilon}^P$
1 (A)	0.438	0.375	10	8	8	12	6.32	5.66	5.66	6.93	6.83
2 (B)	0.313	0.375	10	16	12	12	6.32	8.00	6.93	6.93	6.83
3 (C)	0.188	0.125	20	12	16	18	8.94	6.93	8.00	8.49	8.36
4 (D)	0.063	0.125	20	24	24	18	8.94	9.80	9.80	8.49	8.36
	1	1	60	60	60	60	30.54	30.38	30.38	30.83	30.38

Note: weights are obtained for $\nu = 2$; utilities are obtained for $\varepsilon = 0.5$.

Table 2

Indices obtained for hypothetical population

	$\nu = 2$ $\varepsilon = 0.5$	$\nu = 2$ $\varepsilon = 0$	DAD-DJA $\nu = 2$ $\varepsilon = 0$
$\hat{I}(X_i^x)$	0.188	0.167	0.167
$\hat{I}(N_i^n)$	0.240	0.217	0.217
$\tilde{I}(N_i^x)$	0.210	0.183	0.183
$\hat{I}(N_i^x)$	0.133	0.100	
$\hat{I}(U_{i,\varepsilon}^P)$	0.133	0.100	
$\hat{I}(N_i^E)$	0.108	0.100	0.500
$\hat{\Delta} = \hat{I}(X_i^x) - \hat{I}(N_i^n)$	-0.052	-0.050	-0.050
$\hat{V} = \hat{I}(X_i^x) - \hat{I}(N_i^E)$	0.081	0.067	-0.333
$\hat{C} = \hat{I}(N_i^x) - \hat{I}(N_i^E)$	0.026	0.000	
$\hat{R} = \hat{I}(N_i^n) - \hat{I}(N_i^x)$	0.107	0.117	
$\tilde{C} = \tilde{I}(N_i^x) - \hat{I}(N_i^E)$	0.102	0.083	-0.317
$\tilde{R} = \hat{I}(N_i^n) - \tilde{I}(N_i^x)$	0.030	0.033	0.033

Table 3
Decomposition of redistributive effect for $\nu = 2$ and $\varepsilon = 0$

	A	B	C	A	B	C
				As percentage of $\hat{I}(X_i^x)$		
$\hat{I}(X_i^x)$	0.506665	0.506665	0.506665	100.00	100.00	100.00
$\hat{I}(N_i^n)$	0.291511	0.291511	0.291511	57.54	57.54	57.54
$\tilde{I}(N_i^x)$			0.244328			48.22
$\hat{I}(N_i^x)$	0.244337	0.244337		48.22	48.22	
$\hat{I}(N_i^E)$	0.244540	0.244517	0.251432	48.26	48.26	49.62
				As percentage of $\hat{\Delta}$		
$\hat{\Delta}$	0.215154	0.215154	0.215154	100.00	100.00	100.00
$\hat{V}$	0.262124	0.262148	0.255233	121.83	121.84	118.63
$\hat{C}$	-0.000203	-0.000180	-0.007104	-0.09	-0.08	-3.30
$\hat{R}$	0.047174	0.047174	0.047183	21.93	21.93	21.93
$\hat{\mu}(N_i^E)/\hat{\mu}(N_i^x)$	0.999732	1.000000	1.017489			

Table 4

Decomposition of redistributive effect for different combinations of ν and ε

|	A	B	C	A	B	C
	$\nu = 1.5$ and $\varepsilon = 0$			As percentage of $\hat{\Delta}$		
$\hat{\Delta}$	0.143358	0.143358	0.143358	100.00	100.00	100.00
$\hat{V}$	0.167811	0.167880	0.159522	117.06	117.11	111.28
$\hat{C}$	-0.000139	-0.000070	-0.008432	-0.10	-0.05	-5.88
$\hat{R}$	0.024592	0.024592	0.024596	17.15	17.15	17.16
	$\nu = 3$ and $\varepsilon = 0$			As percentage of $\hat{\Delta}$		
$\hat{\Delta}$	0.287304	0.287304	0.287304	100.00	100.00	100.00
$\hat{V}$	0.374837	0.374568	0.368302	130.47	130.37	128.19
$\hat{C}$	-0.000258	-0.000527	-0.006815	-0.09	-0.18	-2.37
$\hat{R}$	0.087792	0.087792	0.087813	30.56	30.56	30.56
	$\nu = 1.5$ and $\varepsilon = 0.5$			As percentage of $\hat{\Delta}$		
$\hat{\Delta}$	0.303136	0.303136	0.303136	100.00	100.00	100.00
$\hat{V}$	0.350112	0.350366	0.343239	115.50	115.58	113.23
$\hat{C}$	0.023835	0.024088	0.016965	7.86	7.95	5.60
$\hat{R}$	0.023142	0.023142	0.023138	7.63	7.63	7.63
	$\nu = 2$ and $\varepsilon = 0.5$			As percentage of $\hat{\Delta}$		
$\hat{\Delta}$	0.352147	0.352147	0.352147	100.00	100.00	100.00
$\hat{V}$	0.422453	0.422651	0.416784	119.96	120.02	118.36
$\hat{C}$	0.026372	0.026570	0.020713	7.49	7.55	5.88
$\hat{R}$	0.043934	0.043934	0.043924	12.48	12.48	12.47
	$\nu = 2$ and $\varepsilon = 0.9$			As percentage of $\hat{\Delta}$		
$\hat{\Delta}$	0.577556	0.577556	0.577556	100.00	100.00	100.00
$\hat{V}$	0.670484	0.670828	0.665771	116.09	116.15	115.27
$\hat{C}$	0.051673	0.052018	0.047002	8.95	9.01	8.14
$\hat{R}$	0.041255	0.041255	0.041213	7.14	7.14	7.14
	$\nu = 1$ and $\varepsilon = 0.9$			As percentage of $\hat{\Delta}$		
$\hat{\Delta}$	0.592622	0.592622	0.592622	100.00	100.00	100.00
$\hat{V}$	0.635048	0.635456	0.627648	107.16	107.23	105.91
$\hat{C}$	0.042426	0.042834	0.035025	7.16	7.23	5.91
$\hat{R}$	0.000000	0.000000	0.000000	0.00	0.00	0.00

Figure 1

Scattergram of pre- and post-fiscal incomes
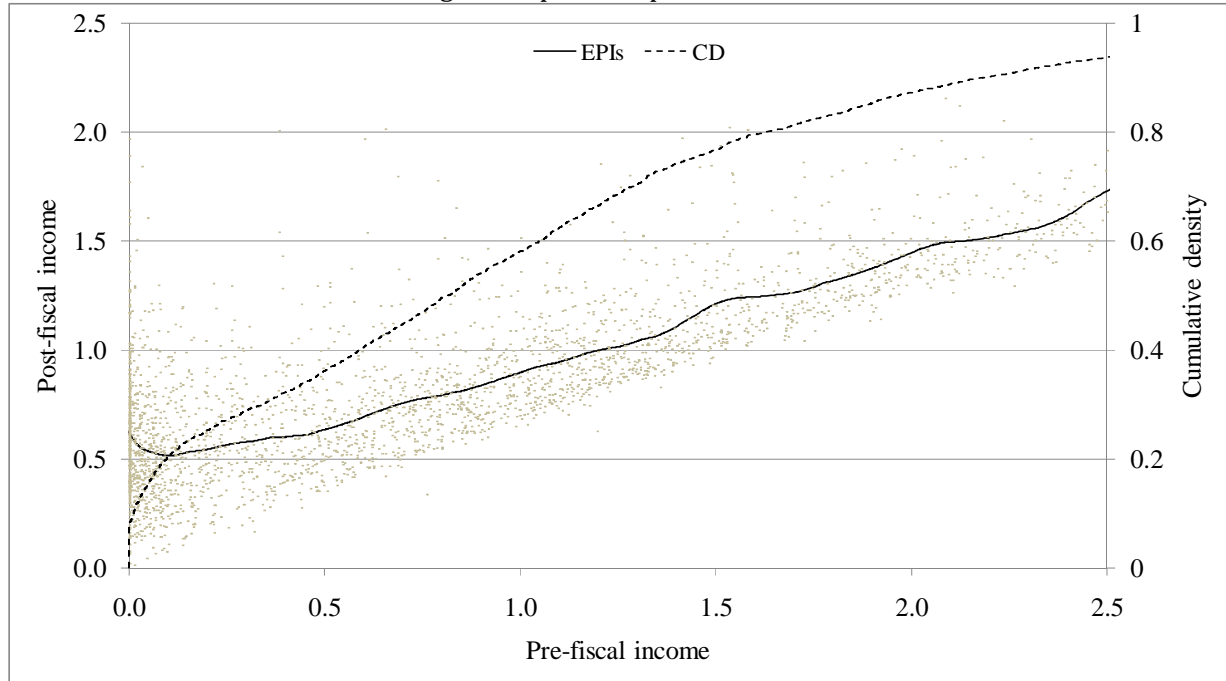
Notes: (a) each point represents one sample income unit; (b) EPIs – expected post-fiscal incomes obtained by KWLPR (see details in section 4.2); (c) CD – cumulative density function.

Figure 2

Scattergram of pre- and post-fiscal incomes
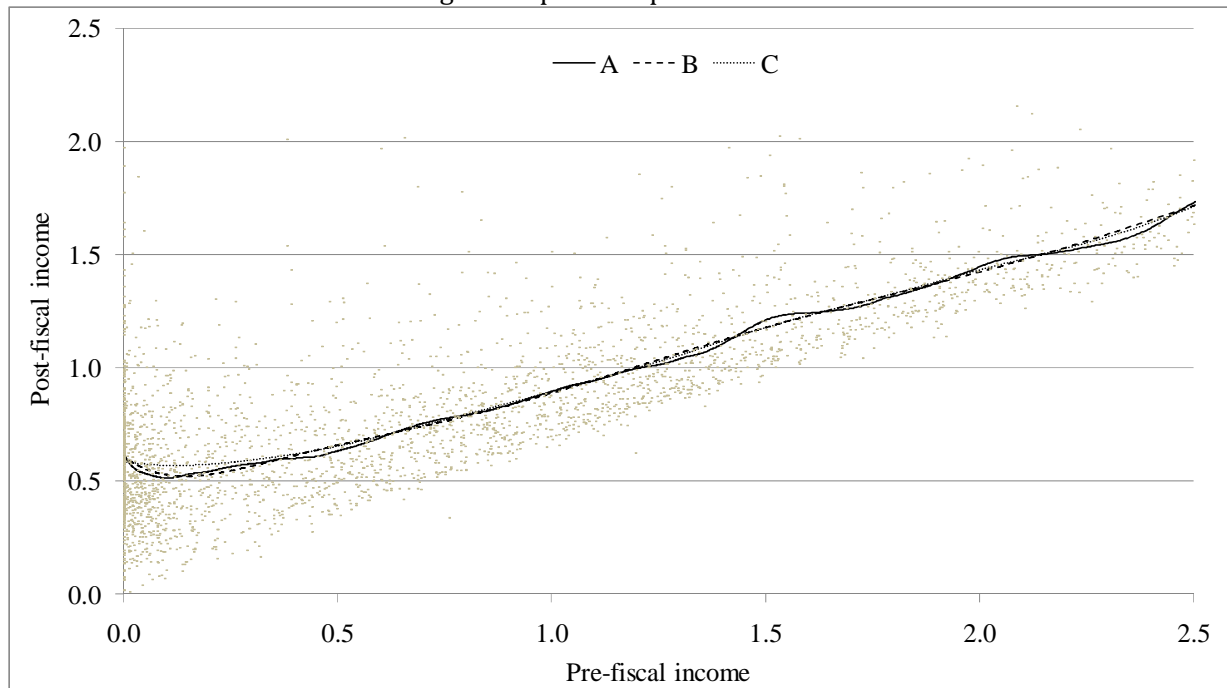
Note: (a) each point represents one sample income unit; (b) A, B, and C – estimates of expected post-fiscal incomes obtained by KWLPR, FTTF, and DJA-LLE fitting methods, respectively (see details in section 4.2).