Distributional Change: Assessing the Contribution of Household Income Sources*

IRYNA KYZYMA,†,‡ ALESSIO FUSCO† and PHILIPPE VAN KERM†,§

†*Luxembourg Institute of Socio-Economic Research (LISER), Esch-sur-Alzette, Luxembourg*
(e-mail: iryna.kyzyma@liser.lu)
‡*IZA, Bonn, Germany*
§*University of Luxembourg, Esch-sur-Alzette, Luxembourg*

Abstract

We develop a decomposition of changes in household income distributions by factor components to quantify the contribution of changes over time in the association between sources of income and changes in their (marginal) distributions. The two components are broken down to isolate the contribution of specific income sources. An application to the change in the distribution of household incomes in Luxembourg between 2004 and 2013 reveals contrasted results: increased association between spouse earnings, public transfers and taxes depressed the income share of poor households while changes in marginal distributions increased incomes in the upper half of the distribution.

I.   Introduction

Inspection of the evolution of income components – earnings, capital income, public or private transfers – is useful to understand changes in the distribution of household incomes. First, distinct components develop differently over time. For example, the literature on the functional distribution of income and related accounts of the changing shares of capital and labour incomes show how the personal distribution of income is shaped by different evolutions of factor prices (e.g. Glyn, 2009; Piketty, 2014; Atkinson and Lakner, 2017; Aaberge, Atkinson and Königs, 2018). Second, households typically receive incomes from multiple sources. The correlation between sources of incomes can mitigate or reinforce inequality. While public transfers, especially if means-tested, normally correlate negatively with market incomes and mitigate inequality in household income (Danziger, Haveman and Plotnick, 1981), the

correlation in the earnings of high-skill, double-income earners tend to reinforce inequality through assortative mating (e.g. Greenwood *et al*., 2014; Eika, Mogstad and Zafar, 2019; Chiappori *et al*., 2020; Frémeaux and Lefranc, 2020). Both the (marginal) distribution of income sources and their association matter to the final distribution of disposable household incomes that policy typically cares about.

Recognizing this fact, Shorrocks (1982) and Lerman and Yitzhaki (1985) have proposed now popular index decomposition approaches to uncover the contribution of income sources to total income inequality.[1] Inequality indices are thereby (additively) decomposed into terms capturing the role of the relative size of each source in total incomes, of inequality in each separate source and of the correlation among sources. Extensions of this method have been used to explain changes in income inequality over time (e.g. Fiorio, 2011; García-Peñalosa and Orgiazzi, 2013; Brewer and Wren-Lewis, 2016). Similar decompositions have been developed to study other income distribution functionals such as indices of poverty (Mussard and Pi Alperin, 2011) or polarization (Deutsch, Fusco and Silber, 2013; Bárcena-Martín, Deutsch and Silber, 2018; Bárcena-Martín and Silber, 2018). Decompositions of summary indices, however, put focus on particular distributional measures, which can make results dependent on the index of interest, for example, the (square of) the coefficient of variation in the case of Shorrocks (1982) or the Gini coefficient for Lerman and Yitzhaki (1985). Furthermore, index decomposition approaches do not allow identifying 'what happens where' in the distribution. For example, a distribution might become more unequal because of the increased dispersion in its upper tail, lower tail or both, and such differences would have different implications for policy response. In this case it is useful to be able to assess the contribution of different sources to, for example, changes in a range of percentiles of the income distribution.

To obtain a more encompassing picture of the factors underlying distributional change over time, researchers have turned to alternative decomposition techniques based on simulation of counterfactual distributions. For example, one simulates the distribution of household incomes that would be observed today if, say, labour incomes had stayed put on past values. Comparisons of actual and such counterfactual distributions allow assessing the impact of changes in sources of income on the entire income distribution, freely from particular summary measures. Burtless (1999), notably, used a rank-preserving income exchange approach to evaluate how changes in the distributions of female and male earnings have contributed to the shift in the distribution of total income in the United States between 1979 and 1996. In a similar spirit, Fournier (2001), Daly and Valletta (2006), Fiorio (2011) and Larrimore (2014) assessed the contributions of changes in various income sources to the inequality trends in Italy, Taiwan and the United States. These studies provide evidence on how shifts in distributions of income sources underlie changes in the distribution of household incomes. We build on this line of research here.

As hinted above, the overall distributive impact of changes in an income source – say, income from capital – depends on two factors: (i) the nature of the change in the

---

[1]See Chantreuil *et al*. (2019) for recent developments.

source distribution itself and (ii) the (change in) association between the source of interest and the remaining household incomes. The first factor refers to the marginal distribution of the source. Is it growing in size? Is it becoming more or less unequally distributed? The second refers to the dependence structure of the various sources of household income. Understanding the contribution of changes in income sources to aggregate distributional change requires sorting out the contribution of these factors. Few studies have attempted to examine systematically how changes in the association between the income sources affected the overall distribution separately from the impact of the change in marginal distributions (see, e.g. Fournier, 2001; Larrimore, 2014). Approaches, however, differ across studies and the literature seems to lack a coherent analytical framework.

The present paper attempts to address this concern by formalizing a general simulation-based, hierarchical decomposition procedure linking simulation approaches to copula theory. In a first step, the change in the distribution of total household incomes is apportioned into two components reflecting (i) changes in the marginal distributions of all income sources and (ii) changes in their dependence structure (as in, e.g. Fournier, 2001; Aaberge *et al.*, 2018). In a second step, the two aggregate components are decomposed further into contributions associated with each income source, separately. This leads to a general decomposition of distributional changes by income source that is similar to those originally presented in Shorrocks (1982) or Lerman and Yitzhaki (1985), but that is not tied to any specific index. Formalizing the decomposition in terms of marginal distributions and explicit copula functions clarifies interpretation of its components and can also guide construction of alternative simulation strategies.[2]

We apply the proposed methodology to the change in the household income distribution in Luxembourg between 2004 and 2013 during which the Gini index increased from 0.266 to 0.303 – a substantial change for a Gini coefficient – while relative income poverty rose by more than 2% points (see e.g. Fusco *et al.* (2014) or Allegrezza and Ametepe (2018)) and Table 1 in section IV. Showing contrasted results along the income distribution, the analysis testifies of the relevance of the decomposition: increased association between spouse earnings, public transfers and taxes depressed the income share of poor households while it is the change in marginal distributions that drove changes in the upper half of the distribution.

The paper is structured as follows. Section II formalizes the decomposition. Section III discusses estimation and implementation. Section IV develops our empirical analysis. Section V concludes. An Online Supplement provides more detailed empirical evidence and sensitivity analyses.

II. Decomposing distributional change by income sources

Our objective is to quantify the contribution of each of K sources of income to the change in an income distribution in a way that allows disentangling the impact arising

---

[2]Copulas have long been used in economics, in particular for welfare comparisons of multidimensional attribute distributions (see, e.g. Atkinson and Bourguignon, 1982; Dardanoni and Lambert, 2001; Abul Naga and Geoffard, 2006; Decancq, 2014; Atkinson and Lakner, 2017; Aaberge *et al.*, 2018).

from a change in the distribution of each source separately from the impact arising from a change in the association across sources.

To do so, we first express the distribution of total income as a function of the joint distribution of income sources using elementary theory for sums of random variables (as, e.g., Fournier (2001) does). We then invoke Sklar's theorem (Sklar, 1959) to reformulate the joint distribution as a function of marginal distributions and of a copula capturing the dependence across sources. Analysing distributions at two points in time, we then define a set of counterfactual distributions obtained by holding subsets of the components of the model constant over time – marginal distributions or components of the copula – and use those to derive a family of additive decompositions of the total distributional change over time.

The distribution of household income and its sources

Consider two cross-sections of a population of N households observed in two time points, 0 and 1. At time $t \in \{0,1\}$, each household $i \in \{1,\ldots, N\}$ receives income from $K > 1$ different sources (e.g. earnings, capital income, public transfers, etc.) so that total income of a household i at time t is

$$y_i^t = \sum_{j=1}^{K} y_{ij}^t \tag{1}$$

where y_{ij}^t is income of household i from source $j \in \{1,\ldots, K\}$.[3]

We view total income as a random variable Y with cumulative distribution function (CDF) $F^t : R \mapsto [0,1]$ and we aim to examine the contribution of the K sources to the change in the distribution function from F^0 to F^1. We treat income sources as K correlated random variables $Y_1, \ldots, Y_K$ with marginal cumulative distribution functions $F_j^t : R \mapsto [0, 1]$ and with joint distribution function $G^t : R^K \mapsto [0, 1]$. Total income is then the sum of K random variables and its cumulative distribution function can be expressed as a function of G^t:

$$\begin{aligned} F^t(y) &= \Pr[Y_1 + \ldots + Y_K \leq y | t] \\ &= \int_{-\infty}^{y} \int_{-\infty}^{y-y_1} \ldots \int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dG^t(y_1, \ldots, y_K). \end{aligned} \tag{2}$$

Equation (2) simply integrates the joint density of all income sources over combinations that add up to a total income less than or equal to y.

Invoking Sklar's theorem (Sklar, 1959) we now express G^t as a function of the K marginal (univariate) CDFs and a copula $C_{1,\ldots,K}^t : [0, 1]^K \mapsto [0, 1]$ (Nelsen, 2006):

$$G^t(y_1, \ldots, y_K) = C_{1,\ldots,K}^t(F_1^t(y_1), \ldots, F_K^t(y_K)). \tag{3}$$

The copula is the joint CDF of K uniformly distributed variables, $r_j^t = F_j^t(y_j)$, representing the ranks of households within the marginal distributions of each separate

---

[3]Some of the sources, such as earnings, may be contributed by multiple members of the household. Some sources, such as taxes, may be negative (deducted from household incomes).

income component. Equation (3) expressly depicts the links between the marginal distributions and their association structure (through the copula) in forming G^t. The copula $C_{1,\ldots,K}^t$ satisfying equation (3) is unique for continuous marginal distributions, but not if any marginal distribution has discrete components, as is typically the case for income sources which have a spike at zero. We return to the implication of the non-uniqueness of the copula in section III.

Substituting equation (3) into (2), F^t can therefore be expressed as a function of the copula $C_{1,\ldots,K}^t$ and the K marginal distributions of income sources F_j^t:

$$F^t(y) = \int_{-\infty}^y \int_{-\infty}^{y-y_1} \cdots \int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^t(F_1^t(y_1), \ldots, F_K^t(y_K)). \tag{4}$$

The representation of the distribution of total income given in (4) is the basis of our decomposition approach. It will allow us to simulate counterfactual distributions that reflect changes in F^t arising from marginal distributions only (through variations in any of the F_j^t) or from the dependence structure (through variations in $C_{1,\ldots,K}^t$). We first use this to form a simple two-term decomposition. We then further decompose each term into a detailed $(2K-1)$-term decomposition that quantifies the separate contribution of each income source to both the copula and the marginal distributions components in the aggregate change from F^0 to F^1.

Changes in marginals vs. changes in rank dependence

Expression (4) leads to a straightforward decomposition of the change in the distribution of total incomes from F^0 to F^1 into a term reflecting changes in marginal distributions and a term reflecting changes in the copula.

Proposition 1. For any pair $(c, m) \in \{0, 1\}^2$ with $c + m = 1$, the change in the CDF of total income from period 0 to period 1 can be decomposed as the sum of two terms:

$$\begin{aligned} \Delta F(y) &= F^1(y) - F^0(y) \\ &= \Delta F_M^{(c)}(y) + \Delta F_C^{(m)}(y) \end{aligned} \tag{5}$$

where $\Delta F_M^{(c)}$ quantifies the contribution of changes in the marginal distributions of the income components and $\Delta F_C^{(m)}$ quantifies the contribution of changes in the rank association of these income components. The two terms are defined as

$$\begin{aligned} \Delta F_M^{(c)}(y) &= \underbrace{\int_{-\infty}^y \int_{-\infty}^{y-y_1} \cdots \int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^{(c)}(F_1^1(y_1), \ldots, F_K^1(y_K))}_{\dot{F}_M^{\{(c),1\}}(y)} \\ &\quad - \underbrace{\int_{-\infty}^y \int_{-\infty}^{y-y_1} \cdots \int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^{(c)}(F_1^0(y_1), \ldots, F_K^0(y_K))}_{\dot{F}_M^{\{(c),0\}}(y)} \end{aligned} \tag{6}$$

and

$$
\begin{aligned}
\Delta F_C^{(m)}(y) \;=\; &\underbrace{\int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^1\Big(F_1^{(m)}(y_1),\;\ldots,\;F_K^{(m)}(y_K)\Big)}_{\ddot{F}_C^{\{1,(m)\}}(y)}\\[2mm]
&-\underbrace{\int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^0\Big(F_1^{(m)}(y_1),\;\ldots,\;F_K^{(m)}(y_K)\Big)}_{\ddot{F}_C^{\{0,(m)\}}(y)}
\end{aligned}
\tag{7}
$$

Proof. Proposition 1 is obtained directly from equation (4). Applying (4) to both $t = 0$ and $t = 1$, we then add and subtract to $F^1 - F^0$ the counterfactual term $\dot{F}_M^{\{1,0\}}(y) = \int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^1\big(F_1^0(y_1),\;\ldots,\quad F_K^0(y_K)\big)$ for $(c,\,m) = (1,\,0)$ or $\dot{F}_C^{\{0,1\}}(y) = \int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^0\big(F_1^1(y_1),\;\ldots,\;F_K^1(y_K)\big)$ for $(c,\,m) = (0,\,1)$ to form (5), (6) and (7) by combining terms. Accordingly, all components except the copula are held fixed to period $m \in \{0,\,1\}$ to form $\Delta F_C^{(m)}(y)$ and all components except the marginal distributions are held fixed to period $c \in \{0,\,1\}$ to form $\Delta F_M^{(c)}(y)$. $\square$

Splitting the total change in the income distribution into two terms as per Proposition 1 isolates changes in the income distribution that come about from a reallocation of incomes from different sources across households, holding the size and distribution of sources constant, from the change in the total income distribution that arises from the separate changes in the distribution of sources themselves. The two terms are reminiscent of the distinction between structural and exchange mobility components in the social and income mobility measurement literature (Markandya, 1982; Van Kerm, 2004).

Proposition 1 effectively defines two alternative decompositions corresponding to $(c,\,m) = (0,\,1)$ and $(c,\,m) = (1,\,0)$. The condition $(c,\;m) \in \{0,\;1\}^2$ imposes that the reference values for the copula and marginal distributions in the counterfactuals are those observed in either period 0 or period 1. The condition $c + m = 1$ is necessary for additive decomposability of the total change: it imposes that if the change in the copula is assessed holding marginal distributions at period 0 then the change in the marginal distributions is assessed holding the copula at period 1, or vice versa.[4] We return to the choice of $(c,\,m)$ in section III.

Detailed decompositions by income sources

The aggregate decomposition given in Proposition 1 is not novel. Although it is expressed differently, it corresponds to the approach adopted in Fournier (2001) or Aaberge *et al.* (2018) (who consider only two sources). We started off, however, aiming

---

[4]Relaxing this restriction and maintaining additivity is possible but implies a third 'interaction' term whose interpretation is not immediate; see Biewen (2014).

to isolate the contribution of distinct income sources to total income distribution changes – in the spirit of index decomposition techniques à la Shorrocks (1982) or Lerman and Yitzhaki (1985) but applied to the entire income distribution. The aggregate decomposition of Proposition 1 provides a general picture of the factors underlying the change in the distribution of household disposable incomes. It, however, raises the question of what specific income source drives these aggregate terms. Is it earnings, or capital incomes, or taxes and transfers (or their association with other sources)?

We therefore further additively decompose each of the two terms of Proposition 1 into distinct contributions from the K sources.

Decomposition of $\Delta F_M^{(c)}(y)$

Ascribing to each of the K sources its distinct contribution to the aggregate marginal distributions component is straightforward.

Proposition 2. For $c \in \{0, 1\}$ and a permutation P of the set of income sources $\{1, 2, \ldots, K\}$, the aggregate contribution of changes in the marginal distributions of income sources to total income change from period 0 to 1 can be additively decomposed as

$$\Delta F_M^{(c)}(y) = \sum_{j=1}^{K} \Delta F_{M,j,P}^{(c)}(y) \tag{8}$$

where $\Delta F_{M,j,P}^{(c)}(y)$ measures the contribution of the change in the marginal distribution of income source j and is defined as

$$\Delta F_{M,j,P}^{(c)}(y) = \underbrace{\int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^{(c)}\left(F_1^{[1 \leq P(j)]}(y_1), \ldots, F_K^{[K \leq P(j)]}(y_K)\right)}_{\ddot{F}_{M,j,P}^{\{(c),1\}}(y)}$$

$$-\underbrace{\int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}} dC_{1,\ldots,K}^{(c)}\left(F_1^{[1 < P(j)]}(y_1), \ldots, F_K^{[K < P(j)]}(y_K)\right)}_{\ddot{F}_{M,j,P}^{\{(c),0\}}(y)}$$

$$\tag{9}$$

where the expressions in square brackets [cond] evaluate to 1 if cond is true and 0 otherwise.

Proof. Proposition 2 is obtained from equation (6) by adding and subtracting $K-1$ counterfactual terms and combining terms to form (9).

The $K-1$ counterfactual terms are constructed from equation (4) by setting $C_{1,\ldots,K}^t$ to $C_{1,\ldots,K}^{(c)}$, the marginal distribution of the first k sources in permutation P to F_j^1 and the remaining sources to F_j^0. This is repeated for the first $k<K$ sources in permutation P to obtain $K-1$ counterfactual terms. $\square$

The contribution of source j in Proposition 2 is measured by the difference between two counterfactual distributions for total income. The first is obtained from expression

(4) with the copula set to period c and the marginal distributions of source j and all sources k appearing earlier in the sequence defined by permutation P, $P(k) < P(j)$, set to 1. The second is obtained similarly but with the marginal distribution of source j set to period 0. The difference between the two is determined by the difference between F_j^0 and F_j^1. This therefore quantifies the contribution of changes over time in source j holding all other components (the copula and the marginal distribution of other sources constant).

Because (4) is nonlinear, the measured contribution of source j is contingent on its position in the permutation P. Proposition 2 therefore defines a family of decompositions indexed by P. The ordering of sources in P is a modelling decision independent on the decision about c and m. We return to decisions about P in section III.

Decomposition of $\Delta F_C^{(m)}(y)$

Quantifying the separate contribution of income sources to the aggregate copula component is more challenging. In order to decompose the copula term in sub-components attributable to changes in the rank dependence between separate income sources, we need to assess the impact of shifting the rank correlation between these income sources from year 0 to 1, one-by-one while keeping other factors unchanged.

To proceed, note how alternative copula-based representations of the joint distribution of sources can be obtained by pooling subsets of sources in expression (4). For example (omitting superscript t for clarity), we can write

$$
\begin{aligned}
G(y_1, \ldots, y_K) &= C_{1,\ldots,K}(F_1(y_1), \ldots, F_K(y_K)) \\
&= C_{(1+2),\ldots,K}(F_{1+2}(y_1 + y_2), F_3(y_3), \ldots, F_K(y_K)) \\
&= C_{(1+2),\ldots,K}\left(\left(\int_{-\infty}^{y_1+y_2} C_{1,2}(F_1(y_1 + y_2 - s), F_2(s))ds\right), F_3(y_3), \ldots, F_K(y_K)\right)
\end{aligned}
$$

$$(10)$$

where the last line is obtained by composing two copula-based expressions, applying expression (4) recursively to the joint distribution of $K-1$ sources obtained after pooling sources 1 and 2 and to the joint distribution of sources 1 and 2. A representation such as expression (10) breaks down the overall K-variate copula term of equation (4) into lower dimensional 'sub-copulas' – an 'outer' copula ($C_{(1+2),\ldots,K}$), and an 'inner' copula ($C_{1,2}$).[5] The K marginal distribution terms remain unchanged. We exploit this procedure to construct counterfactual distributions that capture the impact of changes over time in specific components of the rank dependence structure.

The recursive application of equations (2), (3) and (4) to all different combinations of K sources leads to $2^K - K$ possible representations for G of the form (10). For our decomposition, we only use a subset thereof by adopting a permutation of sources P and considering sequential combination of sources.

---

[5]This is reminiscent but different from hierarchical, vine copula representations (Bedford and Cooke, 2002; Czado, 2019).

To state the results formally, let $\tilde{C}^t_{\Sigma,P,j}:[0,\ 1]^{(K-j+1)}\mapsto[0,\ 1]$ denote the outer copula of the joint distribution of 'outer sources' $\left(\sum_{k=1}^{j}y_{P(k)},\ y_{P(j+1)},\ \ldots,\ y_{P(K)}\right)$ at period t, and $\check{C}^t_{P,j}:[0,\ 1]^j\mapsto[0,\ 1]$ denote the inner copula of the joint distribution of the 'inner sources' $(y_{P(1)}, y_{P(2)},\ldots, y_{P(j)})$. At the boundaries, we have for $j=1$, $\tilde{C}^t_{\Sigma,P,1}(\mathbf{s})=C^t_{1,\ldots,K}(\mathbf{s})$ and $\check{C}^t_{P,1}(s)=s$, and, for $j=K$, $\tilde{C}^t_{\Sigma,P,K}(s)=s$ and $\check{C}^t_{P,K}(\mathbf{s})=C^t_{1,\ldots,K}(\mathbf{s})$. Define finally $\check{F}^{(m)}_{P,j}:R\mapsto[0,\ 1]$, as a counterfactual joint distribution of the inner sources,

$$\check{F}^{(m)}_{P,j}(y)=\int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{j-1}}d\check{C}^1_{P,j}(F^{(m)}_1(y_1),\ \ldots,\ F^{(m)}_j(y_j)). \qquad (11)$$

Proposition 3. For $m\in\{0,1\}$ and a permutation P of the set of income sources $\{1, 2, \ldots, K\}$, the aggregate contribution of changes in the rank dependence between income sources to total income change, $\Delta F^{(m)}_C(y)$, can be additively decomposed as

$$\Delta F^{(m)}_C(y)=\sum_{j=2}^{K}\Delta F^{(m)}_{C,j,P}(y) \qquad (12)$$

where $\Delta F^{(m)}_{C,j,P}(y)$ measures the contribution of the change in the rank order association between income source j and the cumulative total of sources $k<P(j)$:

$$\Delta F^{(m)}_{C,j,P}(y)\quad=\underbrace{\int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}}d\tilde{C}^0_{\Sigma,P,j}\left(\check{F}^{(m)}_{P,j}(y),\ F^{(m)}_{P(j+1)}(y_{j+1}),\ \ldots,\ F^{(m)}_{P(K)}(y_K)\right)}_{\dot{F}^{\{1,(m)\}}_{C,j,P}(y)}$$
$$-\underbrace{\int_{-\infty}^{y}\int_{-\infty}^{y-y_1}\cdots\int_{-\infty}^{y-y_1-\cdots-y_{K-1}}d\tilde{C}^0_{\Sigma,P,j-1}\left(\check{F}^{(m)}_{P,j-1}(y),\ F^{(m)}_{P(j)}(y_j),\ \ldots,\ F^{(m)}_{P(K)}(y_K)\right)}_{\dot{F}^{\{0,(m)\}}_{C,j,P}(y)}$$

$$(13)$$

Proof. Proposition 3 is obtained from Equation (7) by adding and subtracting $K-2$ counterfactual terms and combining terms to form (13). The $K-2$ counterfactual terms are constructed by applying the copula composition procedure to equation (5) with the outer copulas set to period 0 form $\tilde{C}^0_{\Sigma,P,j}$ and the inner copulas set to period 1 form $\check{C}^1_{P,j}$. All marginal distributions are held to $F^{(m)}_k$. The $K-2$ terms are obtained by gradually moving source distributions from the outer copula to the inner copula, in a sequence determined by permutation P. The first source in the permutation is discarded since $\tilde{C}^0_{\Sigma,P,1}(\mathbf{s})=C^0_{1,\ldots,K}(\mathbf{s})$ and the first counterfactual is therefore equal to the rightmost term in (7) and disappears by addition and subtraction. $\qquad\square$

Proposition 3 leads to a $(K-1)$-terms decomposition of the overall copula effect. Each term reflects the contribution of the change over time in the association between a given income source j and the (cumulative total of) previous sources in the permutation P. To achieve additivity of the decomposition, the contribution of source j is evaluated with the association among other sources held at period 1 value for the previous sources in the permutation and at period 0 value for all subsequent sources in

the sequence. Again, Proposition 3 therefore defines a family of decompositions contingent on the permutation P – see section III. By construction, no contribution is assigned to the first source in the sequence defined by permutation P. The first source is a 'root source' against which the contribution of other sources is gradually assessed. The decomposition is therefore degenerate when $K = 2$: clearly, with only two sources, there is no way to attribute change in the association with one or the other source.

Decomposition of index functionals

The decompositions in Propositions 1, 2 and 3 are based on a range of counterfactual representations of the distribution of total household income. The counterfactual distributions are constructed by manipulating components of the joint distributions of income sources as per equation (4) and quantifying how the manipulated components contribute to changes in the distribution of total incomes from period 0 to 1.

   While all results are expressed in terms of decompositions of distribution functions only, it is straightforward to extend all decompositions to changes in functionals thereof, such as quantiles, inequality indices, poverty measures (see, e.g. Cowell, 2011) by applying a functional of interest θ to the counterfactual distributions involved in Propositions 1, 2 and 3.

   Corollary 1. For any pair $(c, \ m) \in \{0, \ 1\}^2$ with $c + m = 1$, permutation P of the set of income components $\{1, 2, \ldots, K\}$, and the change between period 0 and 1 in the index functional $\theta(F)$ expressed as $\Delta\theta = \theta(F^1) - \theta(F^0)$:

1. Aggregate decomposition:

   where
$$\Delta\theta = \Delta\theta_M^{(c)} + \Delta\theta_C^{(m)} \tag{14}$$

$$\Delta\theta_M^{(c)} = \theta\left(\ddot{F}_M^{\{(c),1\}}\right) - \theta\left(\ddot{F}_M^{\{(c),0\}}\right) \tag{15}$$

   captures the contribution of changes in marginal distributions and

$$\Delta\theta_C^{(m)} = \theta\left(\ddot{F}_C^{\{1,(m)\}}\right) - \theta\left(\ddot{F}_C^{\{0,(m)\}}\right) \tag{16}$$

   captures the contribution of changes in the rank association of all sources;

2. Decomposition of the marginal distribution component by source:

$$\Delta\theta_M^{(c)} = \sum_{j=1}^{K} \Delta\theta_{M,j,P}^{(c)} \tag{17}$$

   where

$$\Delta\theta_{M,j,P}^{(c)} = \theta\left(\ddot{F}_{M,j,P}^{\{(c),1\}}\right) - \theta\left(\ddot{F}_{M,j,P}^{\{(c),0\}}\right) \tag{18}$$

3. Decomposition of the rank dependence component by source:

$$\Delta\theta_C^{(m)} = \sum_{j=2}^{K} \Delta\theta_{C,j,P}^{(m)} \tag{19}$$

where

$$\Delta\theta_{C,j,P}^{(m)} = \theta\left(\ddot{F}_{C,j,P}^{\{1,(m)\}}\right) - \theta\left(\ddot{F}_{C,j,P}^{\{0,(m)\}}\right). \tag{20}$$

Proof. Corollary 1 derives directly from Propositions 1, 2 and 3 by applying the functional θ to the counterfactual distributions of total household income involved in propositions. □

Corollary 1 shows the usefulness of defining the decomposition at the level of the distribution function directly, as it provides a unified framework to decompose any functional of interest – a key advantage over particular index decomposition methods.

III. Calculation and implementation

Calculation

All counterfactual distributions used in defining the decomposition are easily estimated from micro-data on incomes following a simulation strategy as in Burtless (1999). The procedure is non-parametric, does not require estimation of the copula functions and can be applied to index functionals directly.[6]

We have a dataset on the K income sources of N households at time $t \in \{0, 1\}$: $\left\{(y_{i1}^t, \ldots, y_{iK}^t)\right\}_{i=1}^N$. Determine first the (fractional) rank of household i's income from source j in year t in the dataset as

$$r_{ij}^t = \hat{F}_j^t(y_{ij}^t) = \frac{1}{N+1}\sum_{h=1}^N I\left\{y_{hj}^t \le y_{ij}^t\right\} \tag{21}$$

(see, e.g. Fan and Patton, 2014) and note that, accordingly,

$$y_{ij}^t = \hat{Q}_j^t(r_{ij}^t) \tag{22}$$

where $\hat{Q}_j^t$ is the generalized inverse (i.e. the quantile function) of $\hat{F}_j^t$.[7] The joint distribution of r_{ij}^t over all sources is the copula.

Proposition 1 requires simulation of the counterfactual distribution of total household incomes at times 0 and 1 holding the copula at period c. These distributions (and index functionals of interest) are obtained using standard estimators from simulated household incomes where household i's income at time t is simulated by holding ranks at period c and applying the period t quantile function:

$$\tilde{y}_i^{(c),t} = \sum_{j=1}^K \hat{Q}_j^t\left(r_{ij}^{(c)}\right). \tag{23}$$

---

[6]All the estimations presented in the paper were performed using the software Stata. The codes can be provided upon request.

[7]Weighting can be introduced in equation (21) if sampling weights need to be applied to the data.

Similarly, simulation of incomes holding marginal distributions constant is given by

$$\tilde{y}_i^{t,(m)} = \sum_{j=1}^{K} \hat{Q}_j^{(m)}\left(r_{ij}^t\right). \tag{24}$$

All terms of the decompositions can then be estimated by $\theta\left(\left\{\tilde{y}_i^{(c),0}\right\}_{i=1}^{N}\right)$, $\theta\left(\left\{\tilde{y}_i^{(c),1}\right\}_{i=1}^{N}\right)$, $\theta\left(\left\{\tilde{y}_i^{0,(m)}\right\}_{i=1}^{N}\right)$ and $\theta\left(\left\{\tilde{y}_i^{1,(m)}\right\}_{i=1}^{N}\right)$ where θ denotes a usual index functional estimator (or the empirical CDF estimator for the decomposition of the CDF directly as in Propositions 1, 2 and 3).

For the detailed decomposition of the marginal distributions term, household incomes are simulated by holding fractional ranks to period c and the quantile functions of the first $j-1$ sources in permutation P to 1 and the others to 0

$$\tilde{y}_i^{t,j,(c)} = \sum_{j'=1}^{j-1} \hat{Q}^1_{j'}\left(r_{ij}^{(c)}\right) + \sum_{j'=j}^{K} \hat{Q}^0_{j'}\left(r_{ij'}^{(c)}\right). \tag{25}$$

The counterfactual distributions needed in equation (8) are again derived from the simulated vectors $\left\{\tilde{y}_i^{t,j,(c)}\right\}_{i=1}^{N}$ obtained for all sources.

Finally, the detailed decomposition of the copula term in equation (12) uses household counterfactual incomes defined as

$$\tilde{y}_i^{(m),t,j} = \hat{Q}^1_{1+\ldots+j}\left(r_{i,1+\ldots+j}^{(m)}\right) + \sum_{j'=j+1}^{K} \hat{Q}^0_{j'}\left(r_{ij'}^{(m)}\right) \tag{26}$$

where $\hat{Q}^t_{1+\ldots+j}$ is the period t quantile function of pooled income sources 1 through j and $r_{i,1+\ldots+j}^{(m)}$ is the corresponding fractional rank of household i in period m in the pooled income distribution.

Income ties and the uniqueness of copulas

Application of Sklar's theorem in equation (3) determines a unique copula only when the marginal distributions are continuous; there is then a one-to-one relationship between the copula C and the multivariate distribution G. When at least one of the marginal distributions has discrete components (e.g. several individuals have the same value of capital income), the copula is not unique and different copula functions are consistent with equation (3) (see Nelsen, 2006; Genest and Nešlehová, 2007). Household income sources are mainly continuous except for a probability mass at zero when households do not receive a particular source of income; for example, pension incomes or targeted benefits. Identification of a unique decomposition requires additional assumptions. Building upon Rothe (2012, 2015), it is possible to make the copula unique by defining a set of K rank allocator functions which assign a unique 'latent rank' to tied ranks and which are equal to observed ranks on continuous segments of the variables, preserving a monotonic relationship between the latent and observed ranks. We can build the rank allocator function by adding a small contamination ε_{ij}^t to all incomes, $y_{ij}^t + \varepsilon_{ij}^t$, where ε_{ij}^t is set small enough that

$y_{ij}^t + \varepsilon_{ij}^t < y_{i'j}^t$ for all $y_{i'j}^t > y_{ij}^t$ and $\varepsilon_{i'j}^t \neq \varepsilon_{ij}^t$ for all $i \neq i'$. The latent ranks are then given by the empirical CDF of the contaminated income values.

Three specific rank allocator functions are relevant. The first, which we follow in the main part of the empirical analysis, is to apply random assignment of ranks for tied observations, for example, using random contamination terms ε_{ij}^t ('jittering'). This assumes away any correlation between the latent ranks of tied observations and other income sources (or total income). The other two rank allocators rank tied observations in one income source according to their positions in other income sources. In the first case, the rank allocator assigns latent ranks in increasing order of the sum of incomes from all other sources, namely, $y_i^t > y_{i'}^t \Rightarrow \varepsilon_{ij}^t > \varepsilon_{i'j}^t$ for any households i and i' with $y_{ij}^t = y_{i'j}^t$. This rank allocator effectively selects the copula that maximizes the correlation between source j latent ranks and total income ranks. In the second case, the rank allocator assigns unique latent ranks in decreasing order of the sum of incomes from all other sources, namely $y_i^t > y_{i'}^t \Rightarrow \varepsilon_{ij}^t < \varepsilon_{i'j}^t$ for any i and i' with $y_{ij}^t = y_{i'j}^t$. This selects a copula that minimizes the correlation between source j latent ranks and total income ranks.

Whether the choice of a rank allocator has substantive impact on the resulting decomposition is an empirical question. Given the shape of income sources, mostly continuous on the strictly positive half-line but with a mass at zero, sensitivity to the rank allocator will be increasing with the magnitude of the change in the probability mass at zero between period 0 and 1. In our empirical application, results show little sensitivity to the choice of rank allocator, even though our data exhibit some relatively large change over time in the proportion of households with no income from some sources; see Table A-1 in Online Supplement.

Permutation selection

Proposition 1 defines two decompositions. Since the distribution of household income is a nonlinear function of the copula and the marginal distributions, setting (c, m) to $(0, 1)$ or to $(1, 0)$ results in two alternative decompositions. One might opt for a 'copula first' strategy identifying the contribution of the copula before the contribution of the marginal CDFs of income components and therefore choose $(1, 0)$, or the other way around. It is difficult to find justification for preferring one ('copula first') over the other ('marginal first'). This is a common issue in simulation-based decomposition analysis. Estimation of the two is recommended in empirical applications. Combination of the decompositions can also be undertaken using a Shapley value approach (e.g. Sastre and Trannoy, 2002; Shorrocks, 2013).[8] In this two-way setting, a Shapley value approach simply consists in taking the average of the contributions of each term across the two options.

Similarly, Propositions 2 and 3 define two families of decompositions since the resulting contributions are contingent on a permutation P. This leads to 2^{K-1} possibly distinct values for the contribution of a source j across the $K!$ possible permutations. Unlike for the choice of (c, m), specific permutations of income sources may be

---

[8]Chantreuil and Trannoy (2013) and Chantreuil *et al.* (2019) provide in-depth discussions of application of the Shapley value to decompositions of inequality functionals, including decompositions by income components. They combine related simulation approaches and the Shapley procedure, but focus on inequality functionals and on (static) contributions of sources to cross-sectional inequality.

meaningful. For example, one may treat pretax market incomes first, then transfer incomes, and handle taxes last. Such a permutation reflects the fact that transfer incomes and taxes are determined by market incomes (at least partly). It may therefore be relevant to examine the change in market income first (holding taxes and transfers at their period 0 values) and the change in taxes and transfers next (as they follow the change in market incomes). With a 'main breadwinner' model in mind, in which the labour supply and earnings of the second (or third) earner are adjusted in response to the earnings of the primary earner, it may also be justified to treat the primary earners first and the secondary earners second in the permutation. This is, however, application-dependent. Shapley value averaging can also be considered to combine estimates from multiple permutations (Shorrocks, 2013).

One point is important to stress here. Similarly to decision about (c, m), decision about P in the decomposition of effects of marginal distributions is mainly technical. Varying the permutation modifies the fixed values at which the impact of changes from 0 to 1 in each component of interest (the marginal distribution of a source) is measured. This may modify the numerical estimates obtained, but it does not fundamentally changes their interpretation. By contrast, varying the permutation in the decomposition of the copula term is not just technical, it also affects the interpretation of the components. By construction, the components of the copula term capture the impact of changes in the rank association of each source with the sources listed earlier in sequence according to P. Modifying the sequence therefore modifies *what* the decomposition estimates, not just *how* it estimates it. It is therefore important to consider sequences that lead to interpretations of clearest substantive interest. We illustrate this point in empirical results presented in the Online Supplement: we show alternative results obtained by modifying a sequence of sources, specifically by reversing our preferred permutation. The estimates of the impact of marginal distributions are affected very little, but the estimates of the components of the copula decomposition differ more significantly. Examining the results in detail, it becomes clear that the reverse order permutation does not just give different estimates, it also leads to estimates that are of less substantive relevance. Notably, while we can interpret our initial results (shown in section IV) in terms of changes in assortative mating or households' joint labour supply decisions (combining earnings from different household members) or as capturing the regressivity of public transfers or the progressivity of taxes, the reverse permutation does not lead to such relatively well-defined effects.

IV.   Empirical application

The distribution of household disposable income in Luxembourg 2004–2013

We applied our decomposition to a study of the change in the household income distribution in Luxembourg between 2004 and 2013. Data are from the *Socio-Economic Panel 'Liewen zu Lëtzebuerg'* (EU-SILC/PSELL3), an annual representative survey on income and living conditions of individuals and private households. Total net household income is partitioned into seven components:
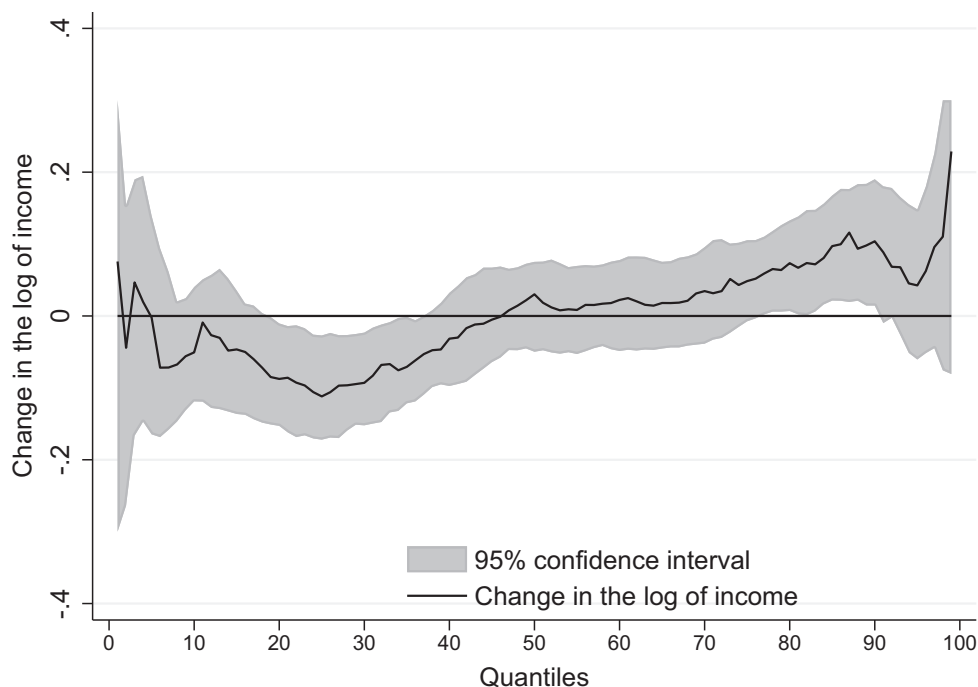
Figure 1. Growth incidence curve: The change in the distribution of household disposable income in Luxembourg between 2004 and 2013
Note: The Figure depicts the differences in the base-2 logarithms of income values between 2013 and 2004, at 99 equally spaced percentiles of the income distribution. The confidence intervals are derived using 500 bootstrap replications. For presentation purposes, we truncated the low bound confidence interval at the first percentile of the income distribution and the upper bound confidence interval at the 98–99th percentiles of the distribution.
Source: Authors' calculations based on the EU-SILC/PSELL III, weighted data (cross-section weights).

$$\text{Total net household income} = Eh + Es + Eo + P + CI + PT - ITC$$

where Eh is gross earnings of the household head, Es gross earnings of spouse, Eo gross earnings of other household members, P pensions, CI capital income, PT public transfers and ITC income taxes and social security contributions. All income components are expressed as single-adult equivalent values using the modified OECD equivalence scale and deflated to 2005 prices.[9] The estimation sample comprises 8,994 individuals for 2004 and 9,963 individuals for 2013.

According to these data, household income inequality increased in Luxembourg between 2004 and 2013 – a period hit by the Great Recession. A growth incidence curve (Figure 1) reveals a relatively simple pattern: incomes at the bottom of the distribution have decreased, those at the top increased. Remarkably, the size of the income growth at the top mirrored the decline at the bottom. Mean income increased faster than median

[9]This equivalence scale gives the value 1 to the first adult, 0.5 to each additional adult and 0.3 to each child below 14. While we follow here standard approach to income distribution analysis, this neglects potential inequality in intra-family allocation of resources (see, e.g. Chiappori, 2016; Chiappori and Meghir, 2016).

TABLE 1

Changes in the distributional summary measures in Luxembourg between 2004 and 2013

	2004	*2013*	*Change between 2004 and 2013*
Mean income	31,345	32,586	+1,241 (675)
Median income	27,680	28,221	+ 540 (655)
P90/P10	3.274	3.642	+0.368 (0.140)
P90/P50	1.801	1.902	+0.101 (0.056)
P50/P10	1.818	1.914	+0.096 (0.058)
Gini	0.266	0.303	+0.037 (0.010)
Poverty rate (%)	13.54	15.72	+2.18 (1.360)
Richness rate (%)	6.92	8.18	+1.26 (0.940)

Note: All measures are obtained for household equivalized disposable income. The relative poverty line is defined as 60% of the median household disposable income. The richness threshold is twice the median household disposable income in a given year. Standard errors in parentheses are derived using 500 bootstrap replications.
Source: Authors' calculations based on the EU-SILC/PSELL III, weighted data (cross-section weights).

income, the Gini coefficient and the P90/P10 ratio both rose significantly (Table 1). Relative poverty and richness rates – measuring the share of individuals above low-income and high-income thresholds – also rose (see Table 1 notes for detailed definitions).

Section A of the Online Supplement details the evolution of each of the seven sources composing our total income variable. In sum, between 2004 and 2013, earnings of household heads tended to decline except for the top 5% – a trend that may be explained by the increase in unemployment which rose from 3.7% in 2004 to 6.8% in 2013. Yet, earnings of spouses, and other household members to a smaller extent, increased. Capital incomes also increased, except at the top, and household incomes were depressed by an increase in taxes, but pensions and other public transfers increased. The magnitudes of the changes and the share of households affected by them vary greatly across sources.

Decomposition analysis

The contrasted evolution of earnings of heads and spouses, the larger sizes of taxes and transfers (and of their progressivity, as gauged by their correlation with market incomes), and the evolution of capital incomes emerge as the potential drivers of the increase in inequality. But the relative importance of these factors is difficult to assess from examining the marginal distributions or even the correlation structure alone.[10] Our decomposition sheds light on this.

[10]Table A-2 of the Online Supplement shows positive correlation in earnings of spouses – an indication of assortative mating – but both of spouses' earnings are negatively correlated with earnings from other household members. Notably the rank correlation in earnings of spouses increased over time – a potentially inequality increasing evolution. Taxes are positively correlated with incomes (except public transfers) as a consequence of their progressivity. Public transfers are negatively associated with market income sources given their role as social insurance. Capital incomes are positively, and increasingly, associated with pensions as a reflection of the accumulation of assets by older population. Capital income also became more correlated with earnings. However, the negative rank correlation of pensions and capital income with public transfers increased over the period suggesting that households who receive higher incomes from these sources became less likely to receive high amounts of benefits. Such households were also paying higher taxes in 2013 as compared to 2004.

Aggregate decomposition

Figure 2 shows aggregate decomposition results, namely how the change in the rank dependence between income sources (Panel A) and their marginal distributions (Panel B) has reflected on the shift in the distribution of household disposable income in Luxembourg. We set (c, m) to $(0, 1)$: the change in the copula is assessed at 2013 marginal distributions and the change in the marginal distributions is assessed at 2004 dependence. The separate effects (shown as dashed lines in Figure 2) add up to the total change (shown as solid line). We use the 'independence' rank allocator to assign unique latent ranks to tied income sources.[11]

The decline in incomes at the bottom of the distribution (between the 10th and 50th percentiles) is mainly driven by the change in the copula. The relative decline in the bottom half of the distribution reflects an increased tendency of households to accumulate low incomes in multiple sources. By contrast, the shifts in the marginal distributions of income sources reflected predominantly on the upper half of the distribution. Panel B in Figure 2 shows that incomes increased from the 30th percentile onwards and this increase explains the change in the household income distribution from the 50th percentile (from which point the copula contribution disappears). Changes in marginal distributions also contributed to a fall in incomes in the bottom 10%. In other words, poorer households who used to receive lower incomes from some sources but higher incomes from other sources became more likely over time to receive low incomes from all sources. By contrast, incomes of the richer households increased but not necessarily in all income sources simultaneously.

In terms of their impact on inequality, both components were dis-equalizing forces: the copula by reducing incomes in (most of) the bottom half of the household income distribution and the marginals by mostly increasing incomes in its upper half. This is reflected in Table 2 which provides decomposition results for summary measures. Both changes in the marginal distributions of income sources and their dependence structure are associated with the increase in all inequality and poverty measures. About three quarters of the increases in the Gini coefficient and the P90/P10 ratio is ascribed to the shifts in the marginal distributions of income sources, and one quarter is ascribed to the copula. By contrast, the change in the copula accounts for a substantial proportion of the increase in the relative poverty rate. According to the estimates, had the rank association between income sources remained the same as in 2004, we would have observed 63.4% smaller increase in the relative poverty rate in 2013.

Detailed decomposition results

The detailed decomposition results look inside the aggregate components. Detailed decompositions as laid down in Propositions 2 and 3 require selecting a particular permutation (or sequence) of sources. Change in the sources down the sequence are assessed with previous sources in the sequence held at their period 1 levels. Moving from market to disposable income sources is a natural option: we start with market incomes (earnings-related first) – earnings of head, of spouse, of other members,

---

[11]Section B of the Online Supplement provides assessment of the sensitivity of our results to alternative choices. See also section Sensitivity of decomposition results to modelling decisions.
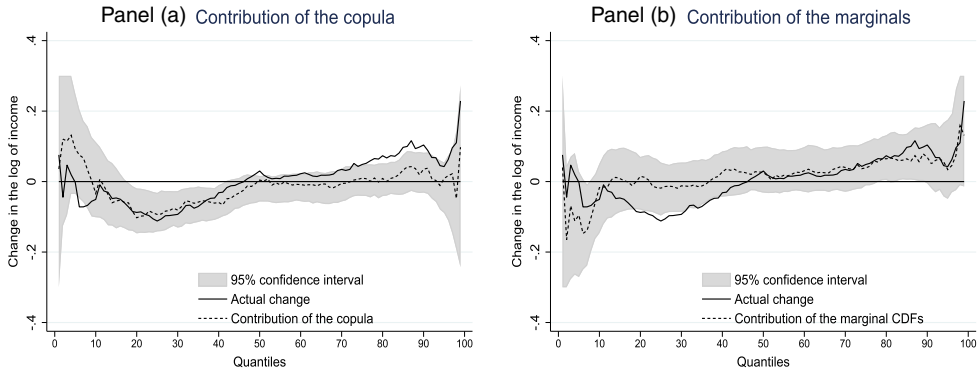
Figure 2. Aggregate decomposition of the change in the household disposable income in Luxembourg between 2004 and 2013
Note: The figure depicts the differences in the logarithms of real income values at 99 equally spaced percentiles of the income distribution. The line 'actual change' captures the difference in actual income values between 2013 and 2004. The contributions of the decomposition components (changes in the copula and marginal cumulative distribution functions of income sources) are calculated according to expressions (6) and (7). The confidence intervals are derived using 500 bootstrap replications.
Source: Authors' calculations based on the EU-SILC/PSELL III, weighted data (cross-section weights).

TABLE 2

Aggregate decomposition of the change in the distribution of household equivalized disposable income

	Overall change	Change attributed to copula		Change attributed to the marginal distributions	
		Absolute	Relative (%)	Absolute	Relative (%)
Gini	+0.029	+0.008	+26.15	+0.021	+73.85
	(0.011)	(0.008)		(0.013)	
P90/P10	+0.362	+0.103	+28.51	+0.259	+71.49
	(0.139)	(0.173)		(0.156)	
P90/P50	+0.099	+0.045	+45.99	+0.053	+54.01
	(0.054)	(0.052)		(0.048)	
P50/P10	+0.095	+0.009	+9.11	+0.087	+90.89
	(0.057)	(0.086)		(0.073)	
Poverty rate (%)	+2.029	+1.286	+63.41	+0.742	+36.59
	(1.30)	(1.31)		(1.07)	
Richness rate (%)	+0.861	+0.058	+6.82	+0.802	+93.18
	(0.89)	(0.78)		(0.81)	

Note: All measures are obtained for household equivalized disposable income. The relative poverty line is defined as 60% of the median household disposable income. The richness threshold is twice the median household disposable income in a given year. Standard errors in parentheses are derived using 500 bootstrap replications.
Source: Authors' calculations based on the EU-SILC/PSELL III, weighted data (cross-section weights).

pensions, capital incomes – and then move to disposable incomes by adding public transfers and finally deducting income taxes paid.[12]

---

[12]We examine and discuss the impact of choosing an alternative permutation in section B of the Online Supplement.

Figure 3 shows detailed results for the contributions of the marginal CDFs to the shift in the distribution of household disposable income. Changes in the marginal distributions of all income sources (with the exception of earnings of other household members and capital income) had a significant contribution to the shift in the distribution of household disposable income. The change in the distribution of earnings of household heads – which declined on average by 8% – resulted in a decline of incomes along the entire distribution of household income (except for the top 10 percentiles), with an especially profound decline documented at the very bottom of the distribution, a potentially strongly inequality-increasing force. The change in the marginal CDF of earnings of spouses – which, by contrast, increased on average by 18% – is associated with a relatively constant, inequality neutral, increase in all percentiles of the total household income distribution. This evidence is in line with findings for other rich countries (see, e.g. Burtless, 1999; Daly and Valletta, 2006; Larrimore, 2014) which show the increased share of spouse's (typically female) earnings in total household income.

Figure 3 also reveals a substantial contribution of pensions to income growth along the entire distribution of household incomes, with larger increases observed higher up the distribution. By contrast to market income sources, changes in the marginal distributions of public transfers and taxes changed in a way to equalize incomes. The change in the marginal distribution of public transfers was associated with the increase in incomes of all households, but especially those at the bottom. Changes in the marginal distribution of taxes had an opposite effect.

Table 3 complements the graphical evidence by quantifying the contributions of changes in the marginal distribution of each income source to distributional summary measures. A positive sign in front of the contribution associated with a particular factor implies that the change in it over time generated an increase in a given summary measure (dis-equalizing effect), and vice versa for a negative sign. As expected from the graphical inspection, the change in the marginal distribution of household heads' earnings was the major driver of the increase in all inequality and poverty measures. The change in the marginal distribution of pensions was the second most important contributor to the increase. These evolutions were partially offset by the shifts in the distributions of public transfers and taxes. Around 50% of the increase in the Gini coefficient induced by the shift in the marginal distribution of earnings of household heads was offset by the changes in the marginal distributions of taxes and transfers.

Figure 4 reveals that it is the change in the rank dependence in market incomes – between earnings of household heads and spouses as well as in the rank dependence of capital income with all earnings-related sources – that has been associated with most of the relative decline in incomes at the bottom of the distribution. Individuals who score low in one market income source became more likely over time to score low also in other income sources, which contributed to the decline in their relative incomes. By contrast, increases in the rank dependence of public transfers and taxes with other income sources – reflecting both higher effective targeting of public transfers and greater tax progressivity – were predominantly beneficial to those at the bottom of the income distribution. These increases, however, did not offset the inequality-increasing impacts of the association of market income sources.
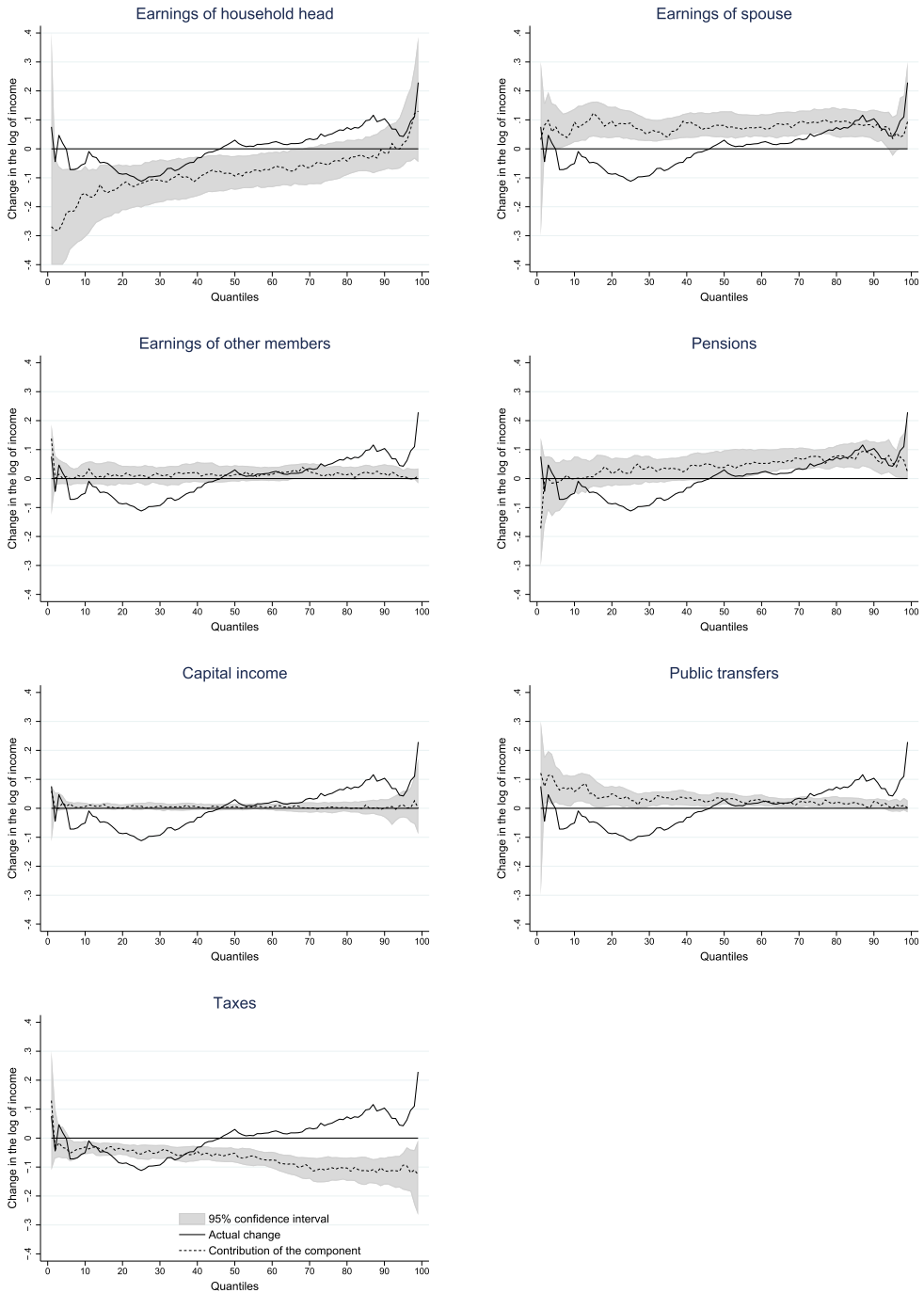
Figure 3. Detailed sequential decomposition of the marginal distributions component
Note: The line 'actual change' captures the actual difference in the logarithm of real income values between 2013 and 2004 at 99 equally spaced percentiles of the income distribution. The confidence intervals are derived using 500 bootstrap replications.
Source: Authors' calculations based on the EU-SILC/PSELL III, weighted data (cross-section weights).

TABLE 3

Detailed decomposition of the change in the distribution of household disposable income

Decomposition component	Gini	P90/P10	P90/P50	P50/P10	Poverty rate	Richness
Overall change	+0.029 (0.011)	+0.362 (0.139)	+0.099 (0.054)	+0.096 (0.057)	+2.03 (1.30)	+0.86 (0.90)
Es with Eh	+0.010 (0.004)	+0.252 (0.095)	+0.063 (0.036)	+0.079 (0.047)	+2.44 (1.01)	+0.66 (0.55)
Eo with Eh, Es	−0.003 (0.002)	−0.050 (0.078)	+0.026 (0.020)	−0.056 (0.047)	−0.99 (1.00)	+0.36 (0.30)
P with Eh, Es, Eo	+0.003 (0.003)	+0.117 (0.096)	−0.006 (0.030)	+0.078 (0.059)	+1.33 (1.28)	+0.10 (0.51)
CI with Eh, Es, Eo, P	+0.007 (0.001)	+0.079 (0.042)	+0.044 (0.022)	+0.003 (0.016)	+0.24 (0.42)	+0.55 (0.35)
PT with Eh, Es, Eo, P, CI	−0.007 (0.002)	−0.115 (0.073)	−0.008 (0.022)	−0.065 (0.043)	−1.59 (1.06)	+0.02 (0.36)
ITC with Eh, Es, Eo, P, CI, PT	−0.002 (0.006)	−0.180 (0.140)	−0.073 (0.044)	−0.030 (0.078)	−0.14 (1.16)	−1.63 (0.74)
Total due to the copula	+0.008 (0.008)	+0.103 (0.173)	+0.046 (0.052)	+0.009 (0.086)	+1.29 (1.31)	+0.06 (0.78)
Eh	+0.030 (0.011)	+0.353 (0.146)	+0.100 (0.036)	+0.093 (0.065)	+1.58 (0.96)	+1.80 (0.69)
Es	−0.001 (0.003)	−0.019 (0.084)	−0.017 (0.032)	+0.006 (0.041)	−0.16 (0.84)	−0.07 (0.49)
Eo	−0.000 (0.001)	+0.019 (0.043)	+0.011 (0.019)	−0.001 (0.019)	+0.05 (0.52)	−0.13 (0.24)
P	+0.007 (0.003)	+0.169 (0.068)	+0.049 (0.033)	+0.047 (0.024)	+0.73 (0.68)	+0.49 (0.40)
CI	−0.000 (0.002)	−0.006 (0.030)	−0.007 (0.014)	+0.004 (0.008)	−0.00 (0.21)	+0.03 (0.26)
PT	−0.006 (0.001)	−0.079 (0.047)	−0.008 (0.012)	−0.037 (0.025)	−1.03 (0.60)	−0.46 (0.002)
ITC	−0.009 (0.005)	−0.178 (0.054)	−0.075 (0.025)	−0.025 (0.018)	−0.43 (0.42)	−0.86 (0.47)
Total due to marginal CDFs	+0.021 (0.013)	+0.259 (0.156)	+0.053 (0.048)	+0.087 (0.073)	+0.74 (1.07)	+0.80 (0.81)

Note: E_h is gross earnings of household head, E_s – gross earnings of spouse, E_o – gross earnings of other household members, P – pensions, CI – capital income, PT – public transfers and ITC – income taxes and social security contributions. All income components are adjusted for the household size using the modified OECD equivalence scale. The standard errors in parentheses are derived using 500 bootstrap replications.
Source: Authors' calculations based on the EU-SILC/PSELL III, weighted data (cross-section weights).
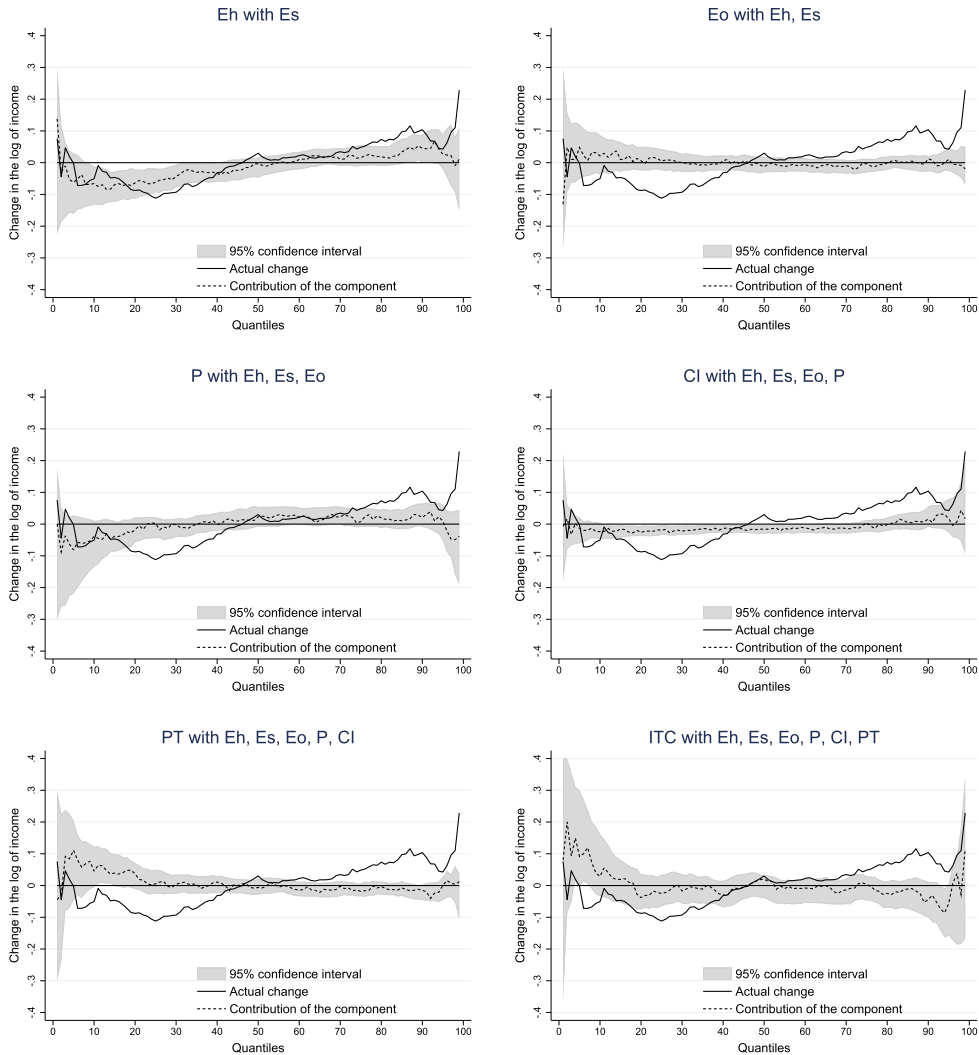
Figure 4. Detailed sequential decomposition of the copula component
Note: The line 'actual change' captures the actual difference in the logarithm of real income values between 2013 and 2004 at 99 equally spaced percentiles of the income distribution. The confidence intervals are derived using 500 bootstrap replications.
Source: Authors' calculations based on the EU-SILC/PSELL III, weighted data (cross-section weights).

Table 3 quantifies these contributions on distributional summary measures. Changes in the rank dependence of earnings of household heads with the earnings of spouses, as well as changes in the rank dependence of capital income with other market income sources, have contributed significantly to the increase in inequality and poverty indexes. The increased rank correlation between earnings of heads and spouses is associated with the largest increase in all inequality and poverty measures. Around one third of the total increase in the Gini coefficient and two thirds of the total increase in all percentile ratios can be assigned to the changed dependence

between earnings of household heads and spouses. With respect to the poverty rate, had the copula between earnings of household heads and spouses remained unchanged, the poverty rate would have been lower in 2013 compared to 2004. Although capital income is small in magnitude, the change in its rank correlation with other pretax sources in our data is associated with a statistically significant decline in incomes at the bottom and in the middle of the household income distribution. Almost 25% of the increase in the Gini coefficient can be attributed to the change in the rank dependence between capital income and other earnings-related income sources. However, this increase has been offset by the changes in the rank correlation of these income sources with transfers. Similarly, the change in the rank dependence of taxes with other income sources further mitigated the dis-equalizing contributions of other factors.

Sensitivity of decomposition results to modelling decisions

Alternative results based on different modelling decisions are presented in section B of the Online Supplement. Following the discussion of section III, three types of sensitivity analysis are performed about (1) the choice of rank allocator, that is, the treatment of tied observations in income components, (2) the choice of (c, m), that is, the order of the decomposition between the two main components and (3) the choice of the permutation P for the detailed decomposition.

The choice of a rank allocator does not appear substantively important in our application, with a cautionary note. In general, we find that at the level of the detailed decomposition the estimates based on two alternative rank allocators are very close to the estimates from the main results. At the level of aggregate decomposition, however, assignment of ranks in line with the assumption of maximum inequality yields somewhat higher contributions associated with the change in marginal distributions of income sources and lower contributions associated with the change in copula as compared to the baseline results, whereas the opposite applies for the assumption of minimum inequality. The choice of (c, m) exhibits similar sensitivity pattern as the choice of rank allocator. Substantive conclusions remain largely unaffected by the choice $(0, 1)$ or $(1, 0)$. The choice of P, finally, can be crucial for reasons already explained in section III. For the sake of illustration, we implemented a reversed permutation from our main analysis. Results for the contribution of sources to the aggregate 'change in marginals' component varied little. However, results for the contribution of sources to the aggregate 'change in copula' component are quite different. As mentioned above, this is not surprising and reflects the variation in the meaning of the components under alternative permutations.

We trust our sensitivity analysis shows reassuring results overall. While specific estimates of the decomposition terms vary with alternative implementation decisions, the major substantive lessons drawn from the analysis are largely unaffected. This is reassuring since we look at a distribution that has changed quite significantly over time, yet, of course, this cannot be a general lesson and sensitivity will inevitably be application specific.

V. Conclusion

This paper proposes a simulation-based decomposition of changes in the distribution of household incomes into two main components: (i) changes in the marginal distributions of different sources of income – labour incomes, capital incomes, transfers, taxes – and (ii) changes in the correlation between these sources. Each component is further disaggregated to distinguish the impacts of changes in each of the separate sources of income. While a number of studies have applied similar principles in counterfactual simulation exercises (e.g. Burtless (1999); Fournier (2001); Larrimore (2014); Aaberge *et al*. (2018)), formalizing the decomposition in terms of the joint distribution of income sources, marginal and copula functions clarifies the rationale and interpretation of the simulations and resulting terms, something that has not been done in the literature for disaggregate decompositions to date.

The decomposition is illustrated on the change in the distribution of total household disposable income in Luxembourg between 2004 and 2013. The distribution of household disposable income had become more unequal between those 2 years, exhibiting an increase in inequality and poverty measures. The decomposition reveals that the shift in the distribution has been induced by both changes in the marginal distributions of income sources and their dependence. While changes in the marginal distributions of income sources can be held predominantly responsible for the growth of incomes in the upper part of the income distribution, the change in association between income sources accounts for the decline in incomes in its lower tail, so both factors contribute to increasing inequality but through different channels. The shifts in the marginal distributions of income sources explain most of the increase in inequality measures and the richness rate whereas the change in the copula is associated with a substantial portion of the increase in the relative poverty rate. Predominantly well-off households have benefited from the growth of incomes coming from various sources over time. The households at the bottom of the distribution not only have not enjoyed this growth but also became more likely to rank low in multiple income sources. Increased progressivity of taxes and transfers only partially offsets this increased correlation of earnings-related sources.

Compared to previous findings for Luxembourg, our approach sheds light on the factors lying behind the change in the entire distribution of household income and a variety of its summary measures. Berger, Fusco and Kyzyma (2014), for example, analysed the change in income inequality in Luxembourg between 1987 and 2010 but focused only on the decomposition of the Gini index using Lerman and Yitzhaki (1985) procedure. Fusco *et al*. (2014) analysed the change in inequality using a set of distributional summary measures but did not perform any sort of decomposition to identify the contributions of various factors to those changes, similarly to Allegrezza and Ametepe (2018), who focused their analysis on poverty. One should be cautious in interpreting our findings in the context of the Great Recession. Although we might capture some distributional trends associated with the economic crisis of 2008, we do not specifically focus on the period right before/after it. The proposed methodology, however, might be useful for studying the impact of recessions given that 'a recession sets several complex mechanisms in motion and a large model with interactions

between income components is required to understand the evolution of income distribution during rapidly rising unemployment' (Aaberge *et al.*, 2000: p. 95).

Final Manuscript Received: June 2021

References

Aaberge, R., Atkinson, A. and Königs, S. (2018). 'From classes to copulas: wages, capital, and top incomes', *Journal of Economic Inequality*, Vol. 16, pp. 295–320.

Aaberge, R., Björklund, A., Jäntti, M., Pedersen, P., Smith, N. and Wennemo, T. (2000). 'Unemployment shocks and income distribution: how did the Nordic countries fare during their crises?', *Scandinavian Journal of Economics*, Vol. 102, pp. 77–99.

Abul Naga, R. and Geoffard, P. (2006). 'Decomposition of bivariate inequality indices by attributes', *Economics Letters*, Vol. 90, pp. 362–367.

Allegrezza, S. and Ametepe, F. (2018). 'Croissance: pro-pauvres ou pro-riches?', *Regards: STATEC: Institut National de la Statistique et des Etudes Economiques, Luxembourg*, Vol. 8.

Atkinson, A. B. and Bourguignon, F. (1982). 'The comparison of multi-dimensioned distributions of economic status', *Review of Economic Studies*, Vol. 49, pp. 183–201.

Atkinson, A. B. and Lakner, C. (2017). Capital and Labor: The Factor Income Composition of Top Incomes in the United States, 1962–2006, World Bank Policy Research Working Paper WPS 8268, The World Bank, Washington DC, USA.

Bárcena-Martín, E., Deutsch, J. and Silber, J. (2018). 'On the decomposition of the Foster and Wolfson bi-polarization index by income sources', *Review of Income and Wealth*, Vol. 64, pp. 853–871.

Bárcena-Martín, E. and Silber, J. (2018). 'On the decomposition of the Esteban and Ray index by income sources', *Econometrics*, Vol. 6, pp. 1–9.

Bedford, T. and Cooke, R. M. (2002). 'Vines: a new graphical model for dependent random variables', *The Annals of Statistics*, Vol. 30, pp. 1031–68.

Berger, F., Fusco, A. and Kyzyma, I. (2014). 'Comprendre l'évolution des inégalités de revenu entre 1987 et 2010 au luxembourg', in Marlier, E., Brosius J., Dautel V., Decoville A., Durand F., Gerber P. and Guio A.-C. (eds), *Cohésion sociale et territoriale au Luxembourg. Regards croisés*, Luxembourg: éditions Peter Lang, pp. 53–68.

Biewen, M. (2014). 'A general decomposition formula with interaction effects', *Applied Economics Letters*, Vol. 21, pp. 636–642.

Brewer, M. and Wren-Lewis, L. (2016). 'Accounting for changes in income inequality: decomposition analyses for the UK, 1978–2008', *Oxford Bulletin of Economics and Statistics*, Vol. 78, pp. 289–322.

Burtless, G. (1999). 'Effects of growing wage disparities and changing family composition on the US income distribution', *European Economic Review*, Vol. 43, pp. 853–865.

Chantreuil, F., Courtin, S., Fourrey, K. and Lebon, I. (2019). 'A note on the decomposability of inequality measures', *Social Choice and Welfare*, Vol. 53, pp. 283–298.

Chantreuil, F. and Trannoy, A. (2013). 'Inequality decomposition values: the trade-off between marginality and efficiency', *Journal of Economic Inequality*, Vol. 11, pp. 83–98.

Chiappori, P.-A. (2016). 'Equivalence versus indifference scales', *Economic Journal*, Vol. 126, pp. 523–545.

Chiappori, P.-A., Costa-Dias, M., Crossman, S. and Meghir, C. (2020). 'Changes in assortative matching and inequality in income: evidence for the UK', *Fiscal Studies*, Vol. 41, pp. 39–63.

Chiappori, P.-A. and Meghir, C. (2016). 'Intrahousehold inequality', in Atkinson A. and Bourguignon F. (eds), *The Handbook of Income Distribution, Volume 2*, Amsterdam: North Holland, pp. 1369–1418.

Cowell, F. A. (2011). *Measuring Inequality*, Perspectives in Economic Analysis, 3rd edn. Oxford: Oxford University Press.

Czado, C. (2019). Analyzing Dependent Data with Vine Copulas: A Practical Guide with R, number 222, in Lecture Notes in Statistics, Springer.

Daly, M. C. and Valletta, R. G. (2006). 'Inequality and poverty in United States: the effects of rising dispersion of men's earnings and changing family behaviour', *Economica*, Vol. 73, pp. 75–98.

Danziger, S., Haveman, R. and Plotnick, R. (1981). 'How income transfers affect work, savings, and the income distribution: a critical review', *Journal of Economic Literature*, Vol. 19, pp. 975–1028.

Dardanoni, V. and Lambert, P. (2001). 'Horizontal inequity comparisons', *Social Choice and Welfare*, Vol. 18, pp. 799–816.

Decancq, K. (2014). 'Copula-based measurement of dependence between dimensions of well-being', *Oxford Economic Papers*, Vol. 66, pp. 681–701.

Deutsch, J., Fusco, A. and Silber, J. (2013). 'The BIP trilogy (bipolarization, inequality and polarization): One saga but three different stories', *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 7, pp. 2013–2022.

Eika, L., Mogstad, M. and Zafar, B. (2019). 'Educational assortative mating and household income inequality', *Journal of Political Economy*, Vol. 127, pp. 2795–2835.

Fan, Y. and Patton, A. J. (2014). 'Copulas in econometrics', *Annual Review of Economics*, Vol. 6, pp. 179–200.

Fiorio, C. V. (2011). 'Understanding Italian inequality trends', *Oxford Bulletin of Economics and Statistics*, Vol. 73, pp. 255–275.

Fournier, M. (2001). 'Inequality decomposition by factor component: a "rank correlation" approach illustrated on the Taiwanese case *Recherches Economiques de Louvain*, Vol. 67, pp. 381–403.

Frémeaux, N. and Lefranc, A. (2020). 'Assortative mating and earnings inequality in France', *Review of Income and Wealth*, Vol. 66, pp. 757–783.

Fusco, A., Van Kerm, P., Alieva, A., Bellani, L., Etienne-Robert, F., Guio, A.-C., Kyzyma, I., Leduc, K., Liégeois, P., Pi Alperin, M.-N., Reinstadler, A., Sierminska, E., Sologon, D., Thill, P., Valentova, M. and Voicu, B. (2014). 'Luxembourg: has inequality grown enough to matter?', in Nolan B., Salverda W., Checchi D., Marx I., McKnight A., Tóth I. and van de Werfhorst H. (eds), '*Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries Experiences*', Oxford: Oxford University Press, pp. 437–458.

García-Peñalosa, C. and Orgiazzi, E. (2013). 'Factor components of inequality: a cross-country study', *Review of Income and Wealth*, Vol. 59, pp. 689–727.

Genest, C. and Nešlehová, J. (2007). 'A primer on copulas for count data', *ASTIN Bulletin*, Vol. 37, pp. 475–515.

Glyn, A. (2009). 'Functional distribution and inequality', in Salverda W., Nolan B. and Smeeding T. M. (eds), *The Oxford Handbook of Economic Inequality*, Oxford, UK: Oxford University Press.

Greenwood, J., Guner, N., Kocharkov, G. and Santos, C. (2014). 'Marry your like: assortative mating and income inequality', *American Economic Review*, Vol. 104, pp. 348–353.

Larrimore, J. (2014). 'Accounting for United States household income inequality trends: the changing importance of household structure and male and female labor earnings inequality', *Review of Income and Wealth*, Vol. 60, pp. 683–701.

Lerman, R. and Yitzhaki, S. (1985). 'Income inequality effects by income source: a new approach and applications to the United States', *Review of Economics and Statistics*, Vol. 67, pp. 151–156.

Markandya, A. (1982). 'Intergenerational exchange mobility and economic welfare', *European Economic Review*, Vol. 17, 307–324.

Mussard, S. and Pi Alperin, M. N. (2011). 'Poverty growth in Scandinavian countries: a Sen multi-decomposition', *Economic Modelling*, Vol. 28, pp. 2842–2853.

Nelsen, R. B. (2006). *An Introduction to Copulas. Springer Series in Statistics*. Springer-Verlag, Berlin.

Piketty, T. (2014). *Capital in the Twenty First Century*. Harvard University Press, Cambridge.

Rothe, C. (2012). 'Partial distributional policy effects', *Econometrica*, Vol. 80, pp. 2269–2301.

Rothe, C. (2015). 'Decomposing the composition effect', *Journal of Business and Economic Statistics*, Vol. 33, pp. 323–337.

Sastre, M. and Trannoy, A. (2002). 'Shapley inequality decomposition by factor components: some methodological issues', *Journal of Economics*, Vol. 9, pp. 51–89.

Shorrocks, A. F. (1982). 'Inequality decomposition by factor components', *Econometrica*, Vol. 50, pp. 193–211.

Shorrocks, A. F. (2013). 'Decomposition procedures for distributional analysis: a unified framework based on the shapley value', *Journal of Economic Inequality*, Vol. 11, pp. 99–126.

Sklar, A. (1959). 'Fonctions de répartition à *n* dimensions et leurs marges', *Publications de l'Institut de Statistique de l'Université de Paris*, Vol. 8, pp. 229–231.

Van Kerm, P. (2004). 'What lies behind income mobility? Reranking and distributional change in Belgium, Western Germany and the USA', *Economica*, Vol. 71, pp. 223–239.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix A. The distribution of income sources in EU-SILC / PSELL III, 2004–2013.
Appendix B. Sensitivity of decomposition results to modelling decisions.