the review of income and wealth

Review of Income and Wealth
Series 67, Number 2, June 2021
DOI: 10.1111/roiw.12470

IINEQUALITY BY POPULATION GROUPS AND INCOME SOURCES: ACCOUNTING FOR INEQUALITY CHANGES IN SPAIN DURING THE RECESSION

BY CARLOS GRADÍN*

United Nations University World Institute for Development Economics Research (UNU-WIDER)

I discuss a new approach which decomposes inequality into the contributions of population groups by income sources. I estimate a matrix with rows and columns which indicate different population groups and income sources, respectively, with each element indicating the marginal change in the inequality contribution of a group (as measured by the Recentered Influence Function) when an income source is added and with all contributions adding up to overall inequality. The approach can be used to analyze the contributions of groups and sources to the trend in inequality over time (or between distributions), disentangling the effect of changes in the composition of the population by groups and changes in their income distribution by sources. An empirical application characterizes the distributional change in Spain following the Great Recession, highlighting the disequalizing role of the massive increase in unemployment or the equalizing effect of social protection through different population groups.

JEL Codes: D31, H24, J64

Keywords: decomposition, Great Recession, income sources, inequality, population groups

1. INTRODUCTION

The analysis of inequality requires the combination of measures which quantify the phenomenon and analytical tools that identify the type of distributional changes which drive inequality trends. Among these tools, the analysis of inequality by population groups and by income sources has played an important role in understanding which characteristics of households or individuals (e.g. place of residence, ethnicity, education, labor market attachment, etc.) and which income sources (e.g. earnings, capital income, social benefits, taxes, etc.) most contribute to explaining inequality at a given moment in time and, more importantly, its changes over time and the differences across countries or regions. These simple decompositions have produced a prolific literature investigating the properties of the different approaches (for a review, see Chakravarty, 2009).

The literature has identified the different ways in which overall inequality indices can be aggregated as a function of inequalities between groups and within

Note: This study has been prepared within the UNU-WIDER project "Inequalities—measurement, implications, and influencing change."

*Correspondence to: Carlos Gradín, United Nations University World Institute for Development Economics Research (UNU-WIDER) Katajanokanlaituri 6 B, FI-00160 Helsinki, Finland (gradin@wider.unu.edu).

groups, with additively decomposable measures (characterized by Shorrocks, 1984) becoming the most popular in empirical analyses. More recently, this approach has been extended by attributing to each group its contribution to overall inequality or to each between-group or within-group component (Gradín, 2020). However, the literature has also investigated how different income sources contribute to total inequality depending on how they change inequality after adding that source (i.e. marginal approaches, following Musgrave and Thin, 1948) or based on their association with total income, that is, natural decomposition of indices that can be written as the weighted sum of incomes (e.g. Shorrocks, 1982; Morduch and Sicular, 2002). These analyses have been recently complemented with the use of regression-based techniques that identify the contribution to inequality of several household characteristics at a time, by directly regressing the log of income, and then, using decomposition properties of the variance of logs (Fields, 2003; Yun, 2006), by reweighting (DiNardo *et al.*, 1996), or by linearizing inequality indices using their "recentered influence function" (RIF) (Firpo *et al.*, 2007, 2009). These methods have become quite popular in distributional analysis, producing a decomposition of distributional changes into compositional and structural effects as an extension of the Blinder–Oaxaca method applied to average differentials in income between two distributions. However, there is little integration of these different tools that have been developed in different branches of the literature.

This paper, therefore, proposes a simple approach to analyze inequality that integrates the joint analysis of population groups and income sources, in a way that is consistent with the use of RIF regression-based decompositions. In doing so, it extends the approach in Gradín (2020) and estimates the contribution of a group to inequality through a specific income source based on the marginal change in the RIF contribution of the group when the source is added. This provides a consistent decomposition with several attractive properties. Given that each group contribution can be expressed as the product of the group population share and the average contribution of the group (based on its distribution), this enables the implementation of a Blinder–Oaxaca-type decomposition of the differential in inequality between two distributions into a compositional effect (changes in population sizes) and a structural effect (changes in the distribution of groups by income source). That is, this is an application of the extended Blinder–Oaxaca decomposition proposed by Firpo *et al.* (2007, 2009). This decomposition can also be implemented in two stages, by computing the aggregate decomposition using reweighting and the detailed decomposition using RIF, making sure that distributional contributions are independent of group sizes. Given that the approach is integrated into the RIF-regression framework (with or without reweighting), the decomposition can be estimated conditional on other covariates too. It can also refer to overall inequality and to the between-group or within-groups terms separately (which in the case of additively decomposable indices will add up to the overall level).

The approach is illustrated using an empirical analysis of the changes in income inequality in Spain after the Great Recession. This is a particularly interesting case to study because of the magnitude of changes in overall inequality, employment, and incomes. Indeed, Spain witnessed a large increase in inequality after being severely hit by the Great Recession when the housing bubble burst, triggering a dramatic financial crisis that affected households, businesses, and the public sector. Real per capita gross domestic product (GDP) fell by 10 percent between 2007 and 2014 only reaching the pre-crisis level in 2017, while the fiscal balance fell from a surplus of 1.9 percent of GDP before the recession to a deficit of 11 percent in 2009 (2.5 percent in 2018).[1] This crisis resulted in massive unemployment (the annual rate rose from 8 percent to 26 percent in 2013), particularly affecting the most vulnerable groups, with a limited response from the government which was constrained by Spain being a member of the eurozone. Despite that, expenditure in social protection benefits (excluding old-age and survivor benefits) accounted for 11.3 percent of GDP in 2007. It rose to its highest level of 14.2 percent in 2011, largely pushed by unemployment benefits, before declining again to 11.9 percent in 2016. The increase in inequality only started to be gradually reversed after 2014, in line with the improvement in employment levels (15 percent unemployment in 2018), even though much of the employment was of low quality due to the proliferation of part-time, temporary, and low-paying jobs.

In this scenario, an analysis that combines population groups and income sources can help us to better understand the channels of the distributional changes triggered by the recession and the subsequent recovery. Given that inequality is more commonly measured by pooling disposable income within households, our analysis is done by type of household. Households are classified based on the intensity of employment of their working-age members (from inactive to fully employed households). The analysis considers the various sources of household income, including gross market income, and the redistributive effect of different social benefits (pensions, unemployment, and other) and direct taxes. Information on income makes extensive use of administrative records, which is expected to improve the accuracy of self-reported income data which can be subject to the well-known problem of underestimation of some income sources in survey data.

The paper proceeds as follows. Section 2 explains the methodology, Section 3 describes the data and main definitions, Section 4 presents the results of the contributions by groups and income sources, and Section 5 provides the results of the Blinder–Oaxaca decomposition. Section 6 concludes.

2. Methodology

In this approach, I consider the analysis of inequality jointly by population groups and by income sources. Let us consider a population represented by a $N \times J$ matrix $\boldsymbol{y}$ with rows indicating individuals and columns indicating income sources:

$$\boldsymbol{y} = \begin{bmatrix} y_1^1 & \dots & y_1^J \\ \dots & \dots & \dots \\ y_N^1 & \dots & y_N^J \end{bmatrix},$$

such that y_i^j $(i = 1, \dots, N; j = 1, \dots, J)$ is the value of (disposable) income source j accrued by individual i, after attributing each individual their household income taking into account household size, with $y^j = \sum_i y_i^j$ representing the total income of source j held by the entire population, and $y_i = \sum_j y_i^j$ total income (from all

[1]Data from Eurostat database (Eurostat n.d.).

sources) possessed by individual i. Overall total income in the population is $y=\sum_i y_i=\sum_j y^j=\sum_i \sum_j y_i^j$. I also consider an exhaustive partition of the population made up of $1 \leq K \leq N$ groups. Let $\{k\}$ be the population subset of $1 \leq N_k \leq N$ individuals in group k.

The $K \times J$ inequality *contribution matrix* $\boldsymbol{S}$ is given by:

$$\boldsymbol{S}=\begin{bmatrix} S_1^1 & \dots & S_1^J \\ \dots & \dots & \dots \\ S_K^1 & \dots & S_K^J \end{bmatrix},$$

where S_k^j is the absolute contribution of group k to overall inequality $I(\boldsymbol{y})$ through source j. The corresponding row sum $S_k=\sum_j S_k^j$ is the total inequality contribution of group k, while the column sum $S^j=\sum_k S_k^j$ is the total inequality contribution of source j.

The decomposition is fully consistent if total group and total source sums add up to overall inequality:

$$I(\boldsymbol{y})=\sum_k S_k=\sum_j S^j=\sum_k \sum_j S_k^j.$$

Then, s_k, s^j and s_k^j indicate the corresponding relative group, source, and group by source contributions after dividing the absolute values by $I(\boldsymbol{y})$.

Ideally, the method to obtain S_k^j should be consistent with how the literature has so far obtained S_k and S^j separately, in order to provide an integrated approach that is furthermore in line with partial analyses of sources and groups.

2.1. *Population Groups*

Following Gradín (2020), I estimate S_k as the change in inequality of marginally increasing the relative size of group k. That is, S_k will be the sum of the per capita value of the RIF for all its members:

$$S_k(I(\boldsymbol{y}))=\frac{1}{N}\sum_{i\in\{k\}} RIF\left(y_i, I(\boldsymbol{y})\right).$$

The influence function of an inequality index $IF(x, I(\boldsymbol{y}))$ measures the impact on inequality of marginally increasing the population mass at income x (i.e. like a small contamination) and has an expected value of zero. It was first introduced by Victoria-Feser (1993) to measure the impact of measurement error on inequality indices.

Let $\boldsymbol{y}_\varepsilon$ be a mixture income distribution assigning a probability $1-\varepsilon$ to the original distribution $\boldsymbol{y}$ and ε to the degenerated distribution assigning mass 1 at a point x (Hampel, 1974). Then, the IF is the directional derivative of the inequality index for this mixture distribution when ε goes to zero:

$$\text{IF}\left(x,I\left(\boldsymbol{y}\right)\right)=\frac{\partial}{\partial\varepsilon}I\left(\boldsymbol{y}_{\varepsilon}\right)|_{\varepsilon=0};\quad \text{with}\quad E(\text{IF}\left(x,I\left(\boldsymbol{y}\right)\right)=0.$$

The IF has an expected value of zero and has been interpreted as the impact on the index of a small contamination. In this line, the IF was first introduced in the distributional analysis by M.P. Victoria-Feser and co-authors (Victoria-Feser, 1993; Victoria-Feser and Ronchetti, 1994; Cowell and Victoria-Feser, 1996) to study the robustness of inequality indices to measurement error.

The RIF is obtained after recentering the IF so that the expected value is the target statistic (Firpo *et al.*, 2007):

$$\text{RIF}\left(x;I\left(\boldsymbol{y}\right)\right)=I\left(\boldsymbol{y}\right)+IF\left(x;I\left(\boldsymbol{y}\right)\right);\quad \text{with}\quad E(\text{RIF}\left(x;I\left(\boldsymbol{y}\right)\right)=I\left(\boldsymbol{y}\right).$$

The RIF provides a convenient linearization of any index that, like inequality indices, are non-linear. That is, the index is expressed as the sum of the impact of marginally increasing the weight of each observation (as estimated by their RIF). For example, Firpo *et al.* (2007) have used this feature to obtain a detailed decomposition of changes in non-linear indices (such as quantiles, the variance, or any inequality index) into a compositional and a structural effect as discussed in this paper, in line with the long tradition in labor economics of decomposing changes in linear statistics, such as the mean. For example, Gradín (2016) used the RIF to explain the difference in Gini index between Spain and Germany, and in Spain over time. Gradín (2020) re-interpreted the RIF as the contribution of individuals or population groups to each inequality index, highlighting its decomposability properties.

In order to use the RIF, it is necessary to know the specific expression corresponding to each target index. Fortunately, the expression of all relevant inequality indices is already known. For example, Table 1 provides the expression for the most common indices, Gini, Mean Log Deviation and Theil index. The table also contains the corresponding contribution of each group:

S_k can be obtained from the RIF regressions that have become popular in the analysis of distributional changes, by just regressing the individual RIF values $\text{RIF}\left(y_i;I\left(\boldsymbol{y}\right)\right)$ on group membership dummies ($\lambda_{ik}=1$ if an individual i belongs to group k; 0 otherwise). The group contribution is the conditional expectation of individual RIF values, if the regression includes no intercept this is given by:

$$\begin{aligned}
&E\left(RIF\left(y_i;I\left(\boldsymbol{y}\right)\right)|\lambda_{ik}=1\right)=\frac{N_k}{N}\beta_k=S_k\left(I\left(\boldsymbol{y}\right)\right);\\
&\text{where}\quad \beta_k=\frac{1}{N_k}\sum_{i\in\{k\}}RIF\left(y_i;I\left(\boldsymbol{y}\right)\right);\\
&\text{with}\quad I\left(\boldsymbol{y}\right)=\sum_k\frac{N_k}{N}\beta_k.
\end{aligned}$$

That is, β_k indicates the average RIF contribution of members of group k, and overall inequality is just the weighted average of the coefficients.

The decomposition by groups is consistent (i.e. the contributions of all groups add up to total inequality), is independent of the path in which groups are

TABLE 1
Summary of RIF for Gini and Generalized Entropy Family (I_α)

	Gini	I_a	$M \equiv I_0$	$T \equiv I_1$
Index	$\frac{n+1}{n} - \frac{2}{n}\sum_{i=1}^{n}\left(\frac{n+1-i}{n}\right)\frac{y_i}{\mu}$	$\frac{1}{\alpha(\alpha-1)}\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\mu}\right)^{\alpha}-1\right]$	$\frac{1}{n}\sum_{i=1}^{n}\ln\frac{\mu}{y_i}$	$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\mu}\ln\frac{y_i}{\mu}$
$IF\left(y_i,I\right)$	$2\frac{y_i}{\mu}\left(\frac{i}{n}-\frac{1+G}{2}\right)+2\left(\frac{1-G}{2}-L_i\right)$	$\frac{1}{\alpha(\alpha-1)}\left[\left(\frac{y_i}{\mu}\right)^{\alpha}-1\right]-I_\alpha-\alpha\left(I_\alpha+\frac{1}{\alpha(\alpha-1)}\right)\frac{y_i-\mu}{\mu}$	$\frac{y_i-\mu}{\mu}+\ln\frac{\mu}{y_i}-M$	$\frac{y_i}{\mu}\ln\frac{y_i}{\mu}-\frac{y_i-\mu}{\mu}-\frac{y_i}{\mu}T$
$RIF\left(y_i,I\right)$	$2\frac{y_i}{\mu}\left(\frac{i}{n}-\frac{1+G}{2}\right)+2\left(\frac{1}{2}-L_i\right)$	$\frac{1}{\alpha(\alpha-1)}\left[\left(\frac{y_i}{\mu}\right)^{\alpha}-1\right]-\alpha\left(I_\alpha+\frac{1}{\alpha(\alpha-1)}\right)\frac{y_i-\mu}{\mu}$	$\frac{y_i-\mu}{\mu}+\ln\frac{\mu}{y_i}$	$\frac{y_i}{\mu}\ln\frac{y_i}{\mu}-(T+1)\left(\frac{y_i-\mu}{\mu}\right)$
S^k	$\frac{2}{n}\sum_{j=1}^{n^k}\frac{j}{n}\frac{y_j}{\mu}-(1+G)\frac{n^k}{n}\frac{\mu^k}{\mu}+\frac{n^k}{n}-\frac{2}{n}\sum_{j=1}^{n^k}L_j^k$	$\frac{n^k}{n}I_\alpha^k-\alpha\left(I_\alpha+\frac{1}{\alpha(\alpha-1)}\right)\frac{n^k}{n}\frac{\mu^k-\mu}{\mu}$	$\frac{n^k}{n}\left[M^k+\frac{\mu^k-\mu}{\mu}+\ln\frac{\mu}{\mu^k}\right]$	$\frac{n^k}{n}\left[\left(\frac{\mu-\mu^k}{\mu}\right)(T+1)+\frac{\mu^k}{\mu}\ln\frac{\mu^k}{\mu}+\frac{\mu^k}{\mu}T^k\right]$

Note: Incomes sorted from poorest to richest; $L_i=\frac{1}{n}\sum_{t=1}^{i}\frac{y_t}{\mu}; L_j^k=\frac{1}{n}\sum_{t=1}^{j}\frac{y_t^k}{\mu}$.
Source: Author's own construction.

considered and is insensitive to the level of aggregation at which groups have been defined. The contribution of a group is the product between its population size and the per capita impact of income on inequality (reflecting how its members are distributed along income levels). The formula to compute the RIF of most inequality indices is already known (e.g. Monti, 1991; Essama-Nssah and Lambert, 2012; Gradín, 2020). The RIF contribution of a population group to inequality indices has the desirable properties (Gradín, 2020). It is U-shaped, assigning higher contributions to extreme incomes, reflecting the normative properties of the target inequality index. That is, a group with its members at the extreme of the distribution will tend to have larger contributions, as a result of inequality indices verifying the Pigou-Dalton Principle of Transfers.[2] The extent of which contributions are larger at each extreme will depend on the specific inequality index used because each index has specific sensitivity to different parts of the distribution. For a given distribution, the contribution of a group increases linearly with its size.[3] The magnitude of the contribution of a group might be large if the group makes up a substantial part of the population and/or the group is highly concentrated at any or both extremes of the income distribution. The contributions to inequality can be estimated separately for between-group and within-group components and, in the case of additively decomposable indices (i.e. the family of generalized entropy measures), the total contribution of a group is the sum of the contribution of both components. This attractive property allows us to integrate the new approach with the conventional decomposition of inequality by subpopulations.

2.2. *Income Sources*

The inequality contribution of an income source S^j will be obtained as the marginal change in inequality after sequentially adding the income source j, when sources $1, \dots, j-1$ are already present:

$$S^1(I(\boldsymbol{y})) = I\left(\boldsymbol{y}^1\right);$$
$$S^j(I(\boldsymbol{y})) = I\left(\boldsymbol{y}^j\right) - I\left(\boldsymbol{y}^{j-1}\right), j = 2, \dots, J;$$
$$\text{where} \quad \boldsymbol{y}^t = \begin{bmatrix} y_1^1 & \dots & y_1^t \\ \dots & \dots & \dots \\ y_N^1 & \dots & y_N^t \end{bmatrix}.$$

is a $N \times t$ submatrix of $\boldsymbol{y}$ containing the first $1 \le t \le J$ sources (columns).

[2]This also means that when analyzing inequality is important to make sure that extreme observations are not due to measurement error as they will have a disproportional influence on the value of the index.

[3]Gradín (2020) shows that in the case of additively decomposable indices, an alternative to RIF group contributions would be to estimate marginal or Shapley group contributions based on the change in inequality after replacing individual incomes in the group by the population average (removing between-group and within-group inequality), thus relaxing the linearity assumption. This does not make much difference in the case of the log mean deviation; both methods (RIF and marginal/Shapley) are empirically equivalent except for relatively very rich groups, but it might have a larger impact on other members of the generalized entropy family. In the case of indices that are not additively decomposable, like Gini, this alternative approach might be less appealing.

Note that S^1 will then be the level of inequality in the distribution of the first source of income across the entire population (the impact on inequality of adding this primary source of income when there is no income). Then, S^j will indicate how inequality subsequently changes after adding each additional source. It will be zero if relative incomes do not change after adding the source (scale invariance principle of relative inequality indices), and, due to the Pigou–Dalton principle of transfers, it will be positive (negative) if adding the source implies a regressive (progressive) redistribution of relative incomes. In general, the sign and magnitude of the contribution of adding a source depend on the incidence of the new source, whether the source is pro-poor or pro-rich with respect to the previous sources, although, as is well known, one needs to account for potentially large re-rankings (such as when a high pension benefit is added to a household with no market income). This decomposition by sources is consistent (the contributions of all sources add up to total inequality).

Alternatively, we can consider the special case in which source j is the last source. This simple marginal approach is the most popular method in the empirical literature proposed by Musgrave and Thin (1948) and studied in the subsequent literature (e.g. Kakwani, 1977), although it does not provide a consistent decomposition of inequality by income sources (unless some renormalization is applied).[4]

The marginal approach provides an appealing and easy interpretation for policy analysis. However, it is well known that the marginal decomposition by income sources does not verify other attractive decomposition properties. In particular, the decomposition is sensitive to the degree of aggregation of sources. Furthermore, any sequence in which the sources are added is ad hoc. For these reasons, the literature has explored several extensions based on the Shapley decomposition (Chantreuil and Trannoy, 2013; Shorrocks, 2013), that is, the expected marginal contribution when all possible sequences are considered $E\left(S^j\right)$ in its simplest version. Given its sensitivity to the level of aggregation of sources, one can also introduce some hierarchy among income sources with more elaborated versions of the Shapley decomposition (such as the Owen or the Nested-Shapley decompositions, e.g. Sastre and Trannoy, 2002; Charpentier and Mussard, 2011; Chantreuil and Trannoy, 2013).

There is no consensus on what approach does the best job and the results of the Shapley decomposition in some cases might be hard to interpret. For example, to obtain disposable income, any sequence that starts subtracting taxes when there is no income is meaningless, and so is starting with social benefits (which typically are contingent on market income) to then add primary income sources. As a consequence, some Shapley decompositions might produce results that are counterintuitive in terms of the degree of progressivity of an income source. Therefore, for simplicity, the analysis here will focus on the sequential marginal approach described above, adding sources in a sequence that makes economic sense: starting with gross primary income, adding gross social benefits, and finally subtracting taxes. This guarantees an easy interpretation while producing a consistent

[4]Alternatively, one can consider removing within-source inequality instead of removing the source, that is, y_i^j are replaced by the corresponding source average $\overline{y}^j = \frac{1}{N} y^j$, for all i (instead of by 0).

decomposition, although the method can be easily adapted to the other versions of the marginal approach, which have their own advantages and inconveniences.

A different approach to estimate the contribution of an income source, but which I am not following here, is based on the natural decomposition of inequality indices such as variance-type indices and Gini (e.g. Shorrocks, 1982; Morduch and Sicular, 2002). This approach applies to indices that can be written as a weighted sum of individual incomes. The contribution of each source is the weighted sum of the source (with the weights estimated with all sources), thus reflecting the degree of association between each income source and total income. Therefore, the method is less general (does not apply to all indices), does not take account of how the income source affects weights, and is generally based on an absolute interpretation of the redistributive effect of income sources (i.e. a uniformly distributed income source has no impact on inequality).

2.3. *Population Groups and Income Sources*

Combining the procedures used to obtain group and source contributions S_k and S^j, one can obtain the marginal RIF contribution of group k to inequality when source j is added to household income, S_k^j:

$$S_k^1 (I(\boldsymbol{y})) = S_k \left(I\left(\boldsymbol{y}^1\right)\right),$$
$$S_k^j (I(\boldsymbol{y})) = S_k \left(I\left(\boldsymbol{y}^j\right)\right) - S_k \left(I\left(\boldsymbol{y}^{j-1}\right)\right), j = 2, \dots, J.$$

The contribution is the product between the population size of the group and its per capita contribution:

$$S_k^j = \frac{N_k}{N} \hat{\beta}_k^j,$$

where $\hat{\beta}_k^j$ is the average marginal per capita contribution of members in group k to inequality when source j is added:

$$\hat{\beta}_k^1 = \beta_k^1;$$
$$\hat{\beta}_k^j = \frac{1}{N_k} \sum_{i \in \boldsymbol{k}} \left[\mathrm{RIF}\left(\sum_{t=1}^{j} y_i^t, I\left(\boldsymbol{y}^j\right) \right) - \mathrm{RIF}\left(\sum_{t=1}^{j-1} y_i^t, I\left(\boldsymbol{y}^{j-1}\right) \right) \right], j = 2, \dots, J;$$

These per capita contributions can be obtained as the difference between the corresponding coefficients estimated in two RIF regressions on group membership dummies, before and after adding source j.

The decomposition is fully consistent with the partial contributions adding up to total inequality, whether they are expressed by sources, by groups, or by the combination of both, given that $\beta_k = \sum_j \hat{\beta}_k^j$:

$$I(\boldsymbol{y}) = \sum_k S_k (I(\boldsymbol{y})) = \sum_j S^j (I(\boldsymbol{y})) = \sum_k \sum_j S_k^j (I(\boldsymbol{y})), \text{with } S^j = \sum_k S_k^j \text{ and } S_k = \sum_j S_k^j.$$

2.4. *Extended Blinder–Oaxaca Decomposition*

A change in inequality between two distributions $\boldsymbol{y}$ and $\boldsymbol{y'}$ (e.g. initial and final distribution) can be written in terms of per capita group contributions and group population shares as:

$$\Delta I \equiv I\left(\boldsymbol{y}'\right) - I\left(\boldsymbol{y}\right) = \sum_k \left(\frac{N'_k}{N'} \beta'_k - \frac{N_k}{N} \beta_k \right) = \sum_k \left(\frac{N'_k}{N'} \sum_j \hat{\beta}'_{kj} - \frac{N_k}{N} \sum_j \hat{\beta}^j_k \right).$$

This expression can be used to decompose the inequality differential into aggregate and detailed compositional and structural (distributional) effects, à la Blinder–Oaxaca (Blinder, 1973; Oaxaca, 1973). This corresponds to the simplest version of the RIF decomposition described in Firpo *et al.* (2007, 2009) applied to this context. One can add and subtract the level of inequality in a counterfactual distribution $\boldsymbol{y}^+$ that combines group population shares $\frac{N_k}{N}$ associated with $\boldsymbol{y}$ with the per capita group contributions β'_k estimated from $\boldsymbol{y}'$: $I\left(\boldsymbol{y}^+\right) = \sum_k \frac{N_k}{N} \beta'_k$.

After rearranging terms, the aggregate decomposition is given by:[5]

$$\Delta I = CE + DE = \sum_k \left(\frac{N'_k}{N'} - \frac{N_k}{N} \right) \beta'_k + \sum_k \frac{N_k}{N} \left(\beta'_k - \beta_k \right).$$

The first term on the right-hand side indicates the aggregate compositional effect CE driven by changes in population shares while keeping the final average group inequality contributions. The second term DE is usually known as the aggregate structural effect and measures the impact on inequality of changing the per capita group contributions, evaluated with the initial group population shares. In our context, the latter just reflects the impact of changes in group income distributions (keeping group sizes constant), and thus, will be called here the distributional effect.

From the previous expression, we can easily obtain the detailed decomposition as:

$$\Delta S^j_k = S'_{kj} - S^j_k = CE^j_k + DE^j_k;$$

with $CE^j_k = \left(\frac{N'_k}{N'} - \frac{N_k}{N} \right) \hat{\beta}'^j_k$ and $DE^j_k = \frac{N_k}{N} \left(\hat{\beta}'_{kj} - \hat{\beta}^j_k \right)$.

This enables us to produce the corresponding matrices for detailed total effects and their compositional and distributional components:

[5]Note that another counterfactual $I\left(\boldsymbol{y}^{++}\right) = \sum_k \frac{N'_k}{N'} \beta_k$ can also be considered: $\Delta I = \sum_k \left(\frac{N'_k}{N'} - \frac{N_k}{N} \right) \beta_k + \sum_k \frac{N'_k}{N'} \left(\beta'_k - \beta_k \right)$. The difference to the previous decomposition is that the compositional and distributional effects are now evaluated with, respectively, the per capita group contribution β_k and the population share $\frac{N'_k}{N'}$. Alternatively, the decomposition could be based on the pool distribution of $\boldsymbol{y}$ and $\boldsymbol{y}'$.

$$\Delta \boldsymbol{S} \equiv \boldsymbol{S}' - \boldsymbol{S} = \boldsymbol{CE} + \boldsymbol{DE}.$$

These matrices can also be aggregated either by groups (row sum) or by sources (column sum):

$$\Delta S_k = S'_k - S_k = \sum_j \left(S'^j_k - S^j_k \right) = CE_k + DE_k = \left(\frac{N'_k}{N'} - \frac{N_k}{N} \right) \beta'_k + \frac{N_k}{N} \left(\beta'_k - \beta_k \right),$$
$$\Delta S^j = S'^j - S^j = \sum_k \left(S'^j_k - S^j_k \right) = CE^j + DE^j = \sum_k \left(\frac{N'_k}{N'} - \frac{N_k}{N} \right) \hat{\beta}'^j_k + \sum_k \frac{N_k}{N} \left(\hat{\beta}'^j_k - \hat{\beta}^j_k \right).$$

The RIF linear approach described above assumes that the per capita group contributions do not depend on group population sizes, that is, the counterfactual can keep per capita contributions when group sizes are changed. Relaxing this assumption might affect the decomposition because above the per capita contributions of both distributions, β_k and β'_k are computed with a different composition of groups in each case. If part of the difference in per capita contributions is due to the change in population sizes, this should be imputed to the compositional effect, not to the distributional effect. By using reweighting to construct the counterfactual distribution, adapting to our context the procedure in Firpo *et al.* (2007), we can produce a new aggregate decomposition into compositional and distributional effects, CE_R and DE_R, which does not rely on the linear assumption.

For example, we obtain the distribution $\boldsymbol{y}^*$, by reweighting the population in $\boldsymbol{y}'$ by the factor $\frac{N_k}{N} / \frac{N'_k}{N'}$, $k = 1, \dots, K$, so that group shares are now the initial ones, $\frac{N_k}{N}$. Following the RIF decomposition of inequality by groups, inequality in this counterfactual distribution is given by $I(\boldsymbol{y}^*) = \sum_k \frac{N_k}{N} \beta^*_k$. A different Blinder–Oaxaca decomposition can be obtained using this new counterfactual by just adding and subtracting inequality in the reweighted sample. After rearranging terms:

$$\Delta I \equiv CE_R + DE_R = \sum_k \left(\frac{N'_k}{N'} \beta'_k - \frac{N_k}{N} \beta^*_k \right) + \sum_k \frac{N_k}{N} \left(\beta^*_k - \beta_k \right).$$

we can easily obtain the relationship between the effects with and without reweighting using two Blinder–Oaxaca decompositions (between $I\left(\boldsymbol{y}'\right)$ and $I(\boldsymbol{y}^*)$, and between $I(\boldsymbol{y}^*)$ and $I(\boldsymbol{y})$), by adding and subtracting $I\left(\boldsymbol{y}^+\right) = \sum_k \frac{N_k}{N} \beta'_k$ in each case:

$$CE_R = \sum_k \left(\frac{N'_k}{N'} \beta'_k - \frac{N_k}{N} \beta^*_k \right) = \sum_k \left(\frac{N'_k}{N'} - \frac{N_k}{N} \right) \beta'_k + \sum_k \frac{N_k}{N} \left(\beta'_k - \beta^*_k \right) = CE + e.$$
$$DE_R = \sum_k \frac{N_k}{N} \left(\beta^*_k - \beta_k \right) = \sum_k \frac{N_k}{N} \left(\beta'_k - \beta_k \right) - \sum_k \frac{N_k}{N} \left(\beta'_k - \beta^*_k \right) = SE - e.$$

That is, the reweighted compositional effect now accounts for the potential impact of changes in population sizes on per capita contributions if the linear assumption does not hold. The size of this specification error $e=\sum_k \frac{N_k}{N}\left(\beta'_k-\beta^*_k\right)$ must be determined empirically. Only when $e=0$, that is, $\beta^*_k=\beta'_k$, will the reweighted and RIF compositional and distributional effects be identical.[6]

The Blinder–Oaxaca detailed decomposition of the distributional effect is known to suffer from an identification problem associated with the need to omit one category in any set of dummies as well as with the scale used in continuous variables. This is because the intercept will capture the effect associated with all omitted categories as well as with zero values for the continuous covariates (Oaxaca and Ransom, 1999). This is usually dealt with by either normalizing the estimated coefficients or not applying any correction, making sure that the omitted categories (along the scale of continuous variables) make economic sense (the choice preferred by Fortin *et al.*, 2011). This problem does not apply in this case given that only one set of dummies (group membership) is considered as an explanatory variable. Therefore, if we use regressions to estimate the effects, no category needs to be omitted. If we omit one category, the intercept will capture its associated distributional effect (the others can be estimated as the sum of the intercept and the corresponding coefficient).

Note that the approach presented here can be adapted to include other covariates in the RIF regressions, as in standard analysis, to estimate a "conditional" decomposition. In this case, the aforementioned identification problem does apply and if RIF is combined with reweighting, a reweighting error r emerges in the computation of the detailed compositional effect $\boldsymbol{SE_R}$ if the reweighted group sizes do not perfectly match the target ones (while if no other covariates are included, then $r=0$).

3. Data and Definitions

For the empirical analysis I use microdata from the Spanish Living Conditions Survey (*Encuesta de Condiciones de Vida*, ECV), conducted by the national statistical office (*Instituto Nacional de Estadística,* INE; described in INE, 2019) as part of the harmonized European Union Statistics on Income and Living Conditions (EU-SILC, Eurostat). I use surveys conducted between 2008 and 2018 in which the data on income are comparable, making extensive use of administrative records (Social Security and Tax Agency).[7] The sample includes about 13,000 households (35,000 individuals) per year.

I measure inequality of household disposable income per equivalent adult (using the modified OECD scales), and the unit of analysis is the individual. Household disposable income is obtained as gross income from various sources

[6]If the analysis is conditional on other covariates, as in the standard RIF analysis, $DE_R=DE+r$, where r would be the reweighting error if the reweighted group size $\frac{N^*_k}{N}$ is not equal to the target $\frac{N_k}{N}$. In our case, $r=0$.

[7]The use of administrative records to produce more accurate income variables started in 2013, but the method has been applied retrospectively since 2008, the earliest year in which matching both data sources using individual identity numbers was possible.

minus direct taxes and social contributions. First, gross income is obtained by subsequently adding: i) market income (i.e. wages, self-employed income, private pensions, other capital income, children's income, private transfers received net of those paid); ii) old-age and survivor pensions; iii) unemployment benefits; and iv) other social benefits (sickness, disability, education grants, child grants, housing grants, other social exclusion). Finally, disposable income is obtained from gross income after subtracting income and wealth taxes and social contributions. This sequence considers market income as the primary source of income, then adds social benefits (that are conditional on market income and work status), and, finally, subtracts taxes (both market income and social benefits are taxable income).

Social benefits in Spain are either contributory or means-tested, the Spanish social system does not include universal cash benefits. In the previous sequence, we considered the fact that pensions are a mixture between market income (as deferred wages) and social assistance (due to the redistributive component both inter and within generations). As a matter of fact, it is often the case that the redistributive impact of social transfers contemplates two alternative scenarios, with pensions included as market income, and with pensions included as social transfers (e.g. Eurostat or the "Commitment to Equity" project). For that reason, the empirical application evaluates the impact of pensions when they are added after market income and before unemployment benefits. For the sake of robustness, however, I also discuss the results from a Shapley decomposition that averages the contribution of all sequences in which the three categories of social benefits can be added after market income and before subtracting taxes (reported in the Appendix). The latter approach is especially attractive in cases in which a more detailed classification of social benefits is used and/or there is no clear sequence in which they can be introduced. The fact that the Shapley is limited to a subset of sources, allows avoiding the undesirable consequences of averaging over all possible sequences previously discussed.

All monetary amounts are expressed in December 2008 constant euros (using the Consumer Price Index for December of the reference year). Standard errors are estimated by clustering individuals by household.

I classify the population into five groups according to how households performed in the labor market during the calendar year preceding interview (which is also the income reference period), using the information about the type of activity reported every month by all working-age members (aged 16 and over) and not just the household head. The first group refers to households reporting no months of being active in the labor market during the year of reference ("inactive households"), as opposed to those reporting being active for at least 1 month ("active households"). Then, "active households" are further disaggregated, according to the employment intensity reported by their members, into: "low intensity" (employed for only half or fewer of active months); "middle intensity" (employed for more than a half of active months, but only employed full time for a half or less); "high intensity" (employed full time for more than a half of the months, with at least one active month not in full-time employment); and "fully employed" (all active months in full-time employment).[8]

[8]The status in each month is self-determined based on rules described in EU-SILC methodology (Eurostat 2017). Information about monthly activities is missing for about 1 percent of observations of individuals aged 16 or over.

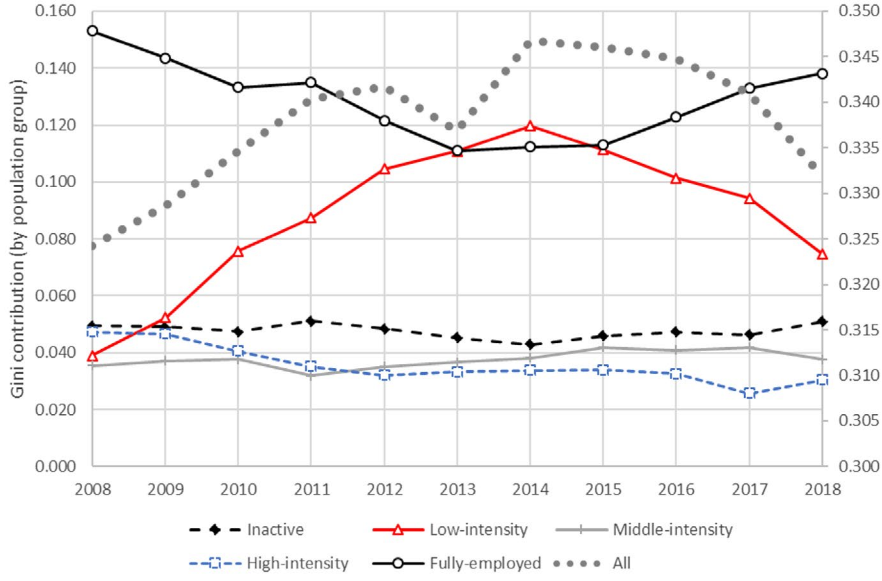

Figure 1. Gini for Disposable Income (All People, Right-Hand Axis) and Decomposition by Population Group (Left-Hand Axis)

Source: Author's construction based on microdata from ECV (INE). [Colour figure can be viewed at wileyonlinelibrary.com]

4. Results: Inequality Decomposition by Income Sources and Population Groups: Total Effects

The impact of the recession on inequality in Spain, especially during the initial years, has already been analysed in relation to different outcomes such as employment (Gradín *et al.*, 2015, 2017), earnings (Bonhomme and Hospido, 2013, 2017), wealth (Amuedo-Dorantes and Borra, 2018), or income (e.g. Gradín, 2016). Already one of the most unequal countries in the European Union (8th out of 27 member states), Spain's disposable income Gini was 0.324 in 2008, following a period of stability with sustained economic growth.[9] The outbreak of the recession, however, raised its level of inequality to a peak of 0.347 in 2014 (0.023 increase), falling back to 0.332 in 2018 (0.015 decline), which was still above the initial level (Figure 1, right-hand axis). The contributions of income sources and groups to inequality are reported in Table 2 for the initial (2008), peak (2014), and latest (2018) years. Table 3 reports the corresponding contributions to inequality changes over time.[10] The results are discussed by sub-period.

4.1. *2008–2014*

When the change in inequality is analyzed by decomposing the contribution of various population groups (left-hand axis of Figure 1 and last column of

[9]Distribution of income at Eurostat database (Eurostat, n.d.). Inequality in Spain had remained stable at around 0.32 between 2005 and 2008 (INE database, n.d.).

[10]Table A1 provides average income values by source over the entire period. Details for the inequality contributions for all years can be found in Table A2 in the Appendix.

TABLE 2
Inequality Contribution Matrix S in 2008, 2014, and 2017 (Gini Index): Population Groups (Rows) and Income Sources (Columns)

	Market Income	Old & Survivor Pensions	Unemployment Benefits	Other Benefits	Direct Taxes	All (S_k)
2008						
Inactive	0.117	−0.056	0.000	−0.005	−0.006	0.049
	0.003	0.002	0.000	0.001	0.000	0.002
Low intensity	0.056	−0.007	−0.005	−0.003	−0.003	0.039
	0.003	0.001	0.000	0.000	0.000	0.002
Middle intensity	0.043	−0.003	−0.001	−0.001	−0.003	0.035
	0.002	0.000	0.001	0.000	0.000	0.002
High intensity	0.057	−0.002	−0.002	−0.002	−0.004	0.047
	0.002	0.000	0.001	0.000	0.000	0.002
Fully employed	0.183	−0.009	−0.003	−0.004	−0.015	0.153
	0.003	0.001	0.000	0.000	0.000	0.003
All (S^j)	0.456	−0.076	−0.010	−0.015	−0.031	0.324
	0.004	0.002	0.001	0.001	0.001	0.004
2014						
Inactive	0.129	−0.074	0.000	−0.006	−0.006	0.043
	0.003	0.002	0.000	0.001	0.000	0.001
Low intensity	0.182	−0.019	−0.025	−0.009	−0.010	0.120
	0.004	0.001	0.001	0.001	0.000	0.003
Middle intensity	0.049	−0.003	−0.003	−0.001	−0.004	0.038
	0.002	0.000	0.000	0.000	0.000	0.002
High intensity	0.040	−0.001	−0.001	−0.001	−0.003	0.034
	0.002	0.000	0.001	0.000	0.000	0.002
Fully employed	0.132	−0.005	−0.001	−0.002	−0.011	0.112
	0.003	0.001	0.000	0.000	0.000	0.003
All (S^j)	0.532	−0.101	−0.030	−0.019	−0.035	0.347
	0.004	0.002	0.001	0.001	0.001	0.003
2018						
Inactive	0.146	−0.081	0.000	−0.007	−0.007	0.051
	0.003	0.002	0.000	0.001	0.000	0.001
Low intensity	0.114	−0.015	−0.010	−0.008	−0.006	0.075
	0.004	0.001	0.001	0.001	0.000	0.003
Middle intensity	0.049	−0.003	−0.002	−0.002	−0.004	0.038
	0.002	0.000	0.000	0.000	0.000	0.002
High intensity	0.037	−0.001	−0.002	−0.001	−0.003	0.030
	0.002	0.000	0.000	0.000	0.000	0.002
Fully employed	0.168	−0.009	−0.002	−0.003	−0.016	0.138
	0.004	0.001	0.000	0.000	0.001	0.003
All (S^j)	0.513	−0.109	−0.015	−0.020	−0.037	0.332
	0.004	0.002	0.001	0.001	0.001	0.004

Note: Bootstraps standard errors below (1,000 replications).
Source: Author's construction based on microdata from ECV (INE).

Tables 1 and 2), it is clear that the increase in inequality during the recession was strongly associated with a larger contribution of the group of households with low-intensity employment (0.081 increase), which was only partially compensated by a smaller contribution from the group of fully employed households or those with high-intensity employment (0.041 and 0.014 reduction, respectively). Indeed, while in 2008 the group of fully employed households was the group that made the largest contribution to inequality (0.153 out of 0.324, i.e. 47 percent), with the other groups contributing 15 percent or less each, the largest contribution in

TABLE 3
Changes Over Time in the Contribution Matrix ($S' - S$), Total Effects (Gini Index): Population Groups (Rows) and Income Sources (Columns)

	Market Income	Old & Survivor Pensions	Unemployment Benefits	Other Benefits	Direct Taxes	All ($S'_k - S_k$)
2008–2014						
Inactive	0.012	−0.018	0.000	0.000	0.000	*−0.007*
	0.004	0.002	0.000	0.001	0.000	0.002
Low intensity	0.126	−0.012	−0.020	−0.006	−0.007	*0.081*
	0.005	0.001	0.001	0.001	0.000	0.004
Middle intensity	0.007	0.000	−0.002	0.000	−0.001	*0.003*
	0.003	0.001	0.001	0.000	0.000	0.003
High intensity	−0.017	0.001	0.001	0.001	0.001	*−0.014*
	0.003	0.000	0.001	0.000	0.000	0.003
Fully employed	−0.052	0.004	0.001	0.002	0.003	*−0.041*
	0.004	0.001	0.001	0.001	0.001	0.004
All (S'^j-S^j)	*0.076*	*−0.025*	*−0.020*	*−0.004*	*−0.004*	*0.023*
	0.006	*0.003*	*0.002*	*0.001*	*0.001*	*0.005*
2014–2018						
Inactive	0.017	−0.007	0.000	−0.001	−0.001	*0.008*
	0.005	0.003	0.000	0.001	0.000	0.002
Low intensity	−0.069	0.004	0.015	0.001	0.003	*−0.045*
	0.006	0.002	0.001	0.001	0.001	0.004
Middle intensity	−0.001	−0.001	0.001	0.000	0.000	*0.000*
	0.003	0.001	0.000	0.000	0.000	0.002
High intensity	−0.003	0.000	−0.001	0.000	0.000	*−0.003*
	0.003	0.000	0.001	0.000	0.000	0.002
Fully employed	0.036	−0.004	−0.001	−0.001	−0.005	*0.026*
	0.004	0.001	0.001	0.000	0.001	0.004
All (S'^j-S^j)	*−0.019*	*−0.007*	*0.015*	*−0.001*	*−0.002*	*−0.015*
	0.006	*0.003*	*0.002*	*0.001*	*0.001*	*0.005*

Note: Bootstraps standard errors below (1,000 replications).
Source: Author's construction based on microdata from ECV (INE).

2014 came from households with low-intensity employment (0.120 or 35 percent, as opposed to 12 percent in 2008), with the contribution of the fully employed households reduced to 32 percent.

Table 2 (last row of each year panel) and Figure 2 show, by income source, that inequality in disposable income in each year was the result of a high level of inequality in market income, which was only partially compensated for by the redistributive impact of social benefits and taxes. The largest redistributive effect is obtained after adding old-age and survivor pensions, followed by the impact of subtracting taxes. When the trend in inequality is analysed by decomposing the contribution of these income sources (Table 3), it is also clear that the inequality increase during the recession was unambiguously associated with a much larger contribution of market income, with a total effect of a 0.076 increase in inequality. This was only partially compensated for by a larger equalizing contribution of pensions and unemployment benefits over the same period, reducing inequality by 0.025 and 0.020). It is interesting to note that despite the large increase in unemployment benefits (which tripled their equalizing effect), the total cushion effect of pensions on the increase in inequality during the recession was larger. The fact that in the presence of multi-generational households old age pensions can play a role

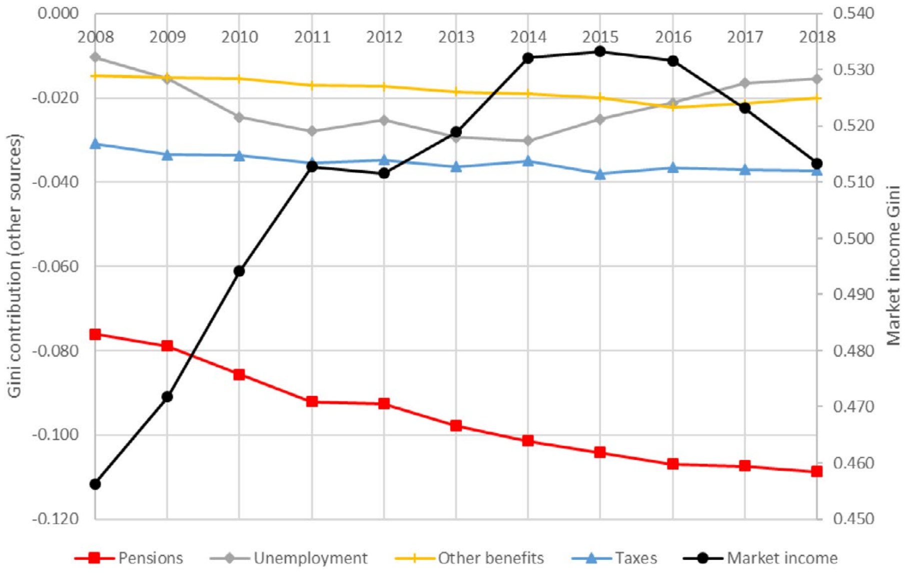

Figure 2. Marginal Inequality Contributions by Income Source (Market Income on the Right-Hand Axis, other Sources on the Left-Hand Axis)

Source: Author's construction based on microdata from ECV (INE). [Colour figure can be viewed at wileyonlinelibrary.com]

different from which they were originally designed for, does not entirely come as a surprise. For example, Diris *et al.* (2017) and Verbist *et al.* (2018) have already highlighted the importance of old age pensions in reducing child poverty in southern European countries, where this type of households is more frequent. Gradín (2016) has already found that the prevalence of extended families in Spain contributes to reducing inequality by diversifying income sources, with retirement pensions playing an important role. Our results suggest that the effect was intensified during the recession, mitigating the increase in inequality.

The approach introduced in Section 2 allows us to integrate both narratives by analyzing in detail the contribution of population groups through the different income sources in a consistent way (Figure 3 and Tables 1 and 2). It becomes clear that inequality increased during the recession due to a much larger contribution of the low-intensity employed group of households to market income inequality (0.126). This large total effect was partially compensated for by different forces going in the opposite direction: i) the reduction in the contributions of highly and fully employed households to market income inequality (0.052 and 0.017 reductions, respectively); ii) the larger equalizing contribution of low-intensity employed households after unemployment benefits were accounted for (0.020 inequality reduction); and iii) the larger equalizing effect associated with inactive and low-intensity employed households through their old-age and survivor pensions (0.018 and 0.012 reductions).[11]

[11]Table A1b in the Appendix shows that the share of old age pensions and unemployment benefits in low-intensity employed households disposable income increased, respectively, from 15.3 and 16.7 percent in 2008 to 17.8 and 22.6 percent in 2014, while the share of market income fell from 71 to 63.9 percent over the same period.

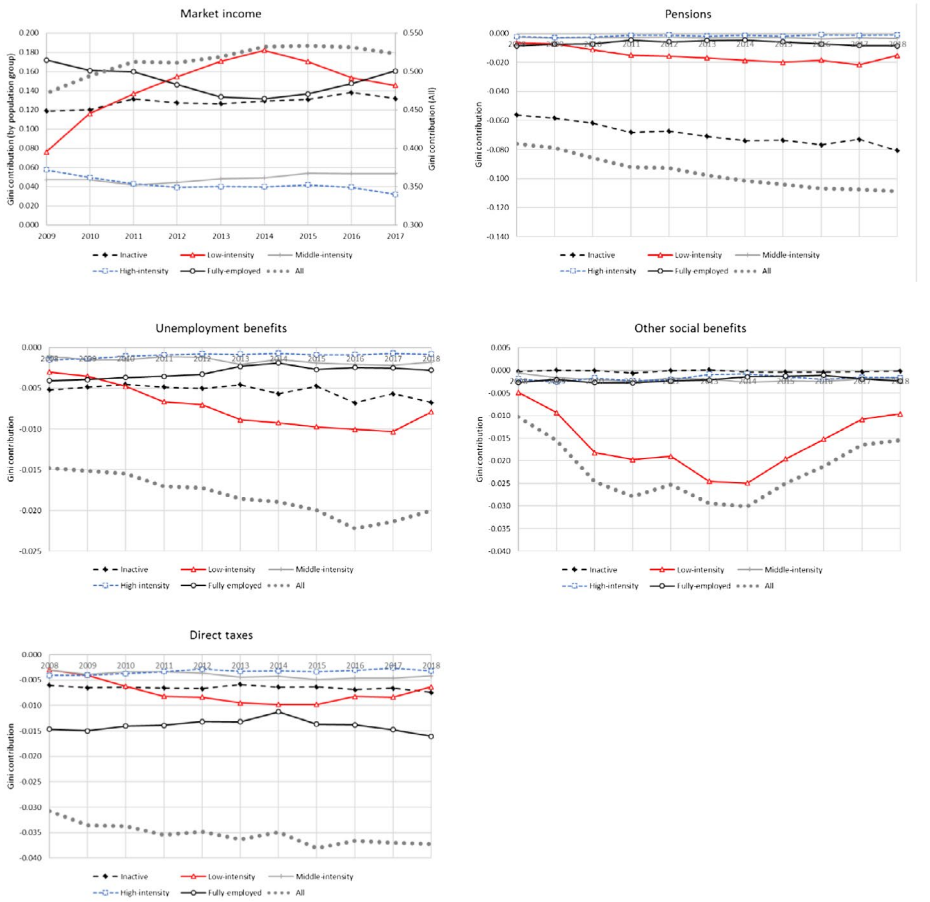

Figure 3. Gini Contribution by Income Source (All People, Right-Hand Axis) and Contribution by Group (Left-Hand Axis)

Source: Author's construction based on microdata from ECV (INE). [Colour figure can be viewed at wileyonlinelibrary.com]

The extent to which these inequality-enhancing or equalizing forces were the result of the radical change in the composition of the population by household type due to massive unemployment, or rather the effect of changes in group income distributions, is analysed in Section 5.

Table A1c in the Appendix explores how sensitive the results are to alternative sequences in which social benefits can be introduced, reporting the Shapley decomposition of the contribution of groups and sources to the decline in inequality. Results are quite robust, with only a slightly smaller contribution of pensions and larger of unemployment benefits channeled through the low-intensity employed group of households, that do not change any of the implications discussed above.

4.2. *2014–2018*

The decline in inequality between 2014 and 2018 was mostly the result of the reversion of the previous trends. On the group side, the fall in inequality was driven by the smaller contribution of the low-intensity employment group of households (0.045 reduction), partially compensated for by the increase in the contribution of fully employed households and, to a lesser extent, inactive households (0.026 and 0.008, respectively). On the income source side, the fall in inequality can be attributed to the declining inequality in market income (0.019 reduction) and, to a lesser extent, the larger equalizing impact of pensions (0.007). The reduction in inequality occurred despite the less equalizing role of unemployment benefits (which explains an increase in inequality of 0.015).

If we again combine the analysis of groups and sources, it becomes clear that households with low-intensity employment account for a reduction in inequality of 0.069 through their lower contribution to market income inequality (reversing 55 percent of the previous increase), partially compensated for by an increase in the corresponding contributions of fully employed and inactive households (0.036 and 0.017). The disequalizing effect of adding unemployment benefits occurs entirely through the low-intensity employment group (0.015), while the equalizing effect when pensions are added is channeled through the contribution of inactive households and through the fully employed group (with a disequalizing effect by the low-intensity group).

The results for this period are also robust to the use of the Shapley decomposition among the social benefits (Table A1c in the Appendix).

5. Results: Disentangling the Compositional and Distributional Effects on Inequality Trends

In this section, I try to disentangle whether the inequality trend over the two sub-periods (2008–2014 and 2014–2018) was the result only of important changes in the composition of the population according to the performance of their households in the labor market, or was also the result of changes in how these groups were distributed along the income distribution. That is, the total effect associated with groups and income sources described in the previous section will be decomposed into their characteristics and structural (distributional) effects.

5.1. *Changes in Group Size and Income Distribution*

There is no doubt that Spain witnessed substantial changes in the composition of households, classified by their involvement in the labor market, after the beginning of the recession, mostly as the result of skyrocketing unemployment rates. These changes may also have been shaped by the recession affecting the demographics of households due to an increase in economic hardship, like reversing the migration flows (from net immigration to net out migration), reducing the emancipation rate of young people, or giving incentives to the reunification of different generations under the same roof.[12]

Figure 4 shows the remarkable increase between 2008 and 2014 in the share of the population in households with low-intensity employment (rising from 9 to 28

[12]However, the net effect was that the average number of members and equivalent adults actually declined, especially during the initial period, from 3.37 and 2.07, respectively, in 2008, to 3.15 and 1.96 in 2014, and 3.12 and 1.95 in 2018.

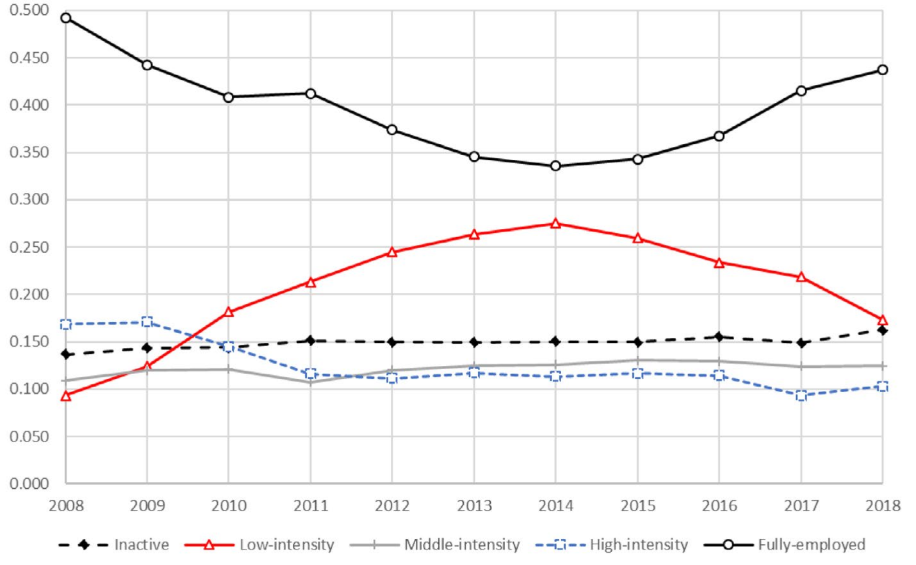

Figure 4. Population Shares by Group ($\frac{N_k}{N}$)

Source: Author's construction based on microdata from ECV (INE). [Colour figure can be viewed at wileyonlinelibrary.com]

percent of the population), alongside the decline in the share of the population in fully employed households and, to a lesser extent, in households with high-intensity employment (respectively from 49 to 34 percent, and from 17 to 11 percent).[13] The trend was partially reversed after 2014. Over the entire period, there was a smoother increase in the share of the population in inactive households, from 14 percent in 2008 to 16 percent in 2018, which is probably the result of an ageing population, aggravated by the reversion of migration flows.

As well as the changes in group sizes, there were also relevant changes in the per capita group contributions to inequality. The per capita contribution of a group reflects how its members (regardless of the number) are distributed along income levels. For example, Figure 5a displays the densities of the different groups in 2018. It is not surprising that the per capita contribution of households is largest among those with low-intensity employment, as these tend to be disproportionally located at the bottom of the disposable income distribution. Inactive households are also disproportionally represented at the bottom, while fully employed households or those with high-intensity employment tend to be over-represented at the top. Thus, changes over time in per capita contributions reflect changes in these distributional patterns. In general, a decline over time in the per capita contribution of a group can be interpreted as the group getting closer to the population mean, while an

[13]Gradín (2016) covered part of the period of increasing inequality, 2009–2012, and identified an important compositional effect in driving the trend of increasing inequality among active households (while there was little change among inactive households), using a standard RIF regression-based decomposition controlling for a number of households characteristics. The analysis presented here extends the period of reference and enables a more integrated analysis of income sources and population groups for the whole population while remaining in the RIF framework.

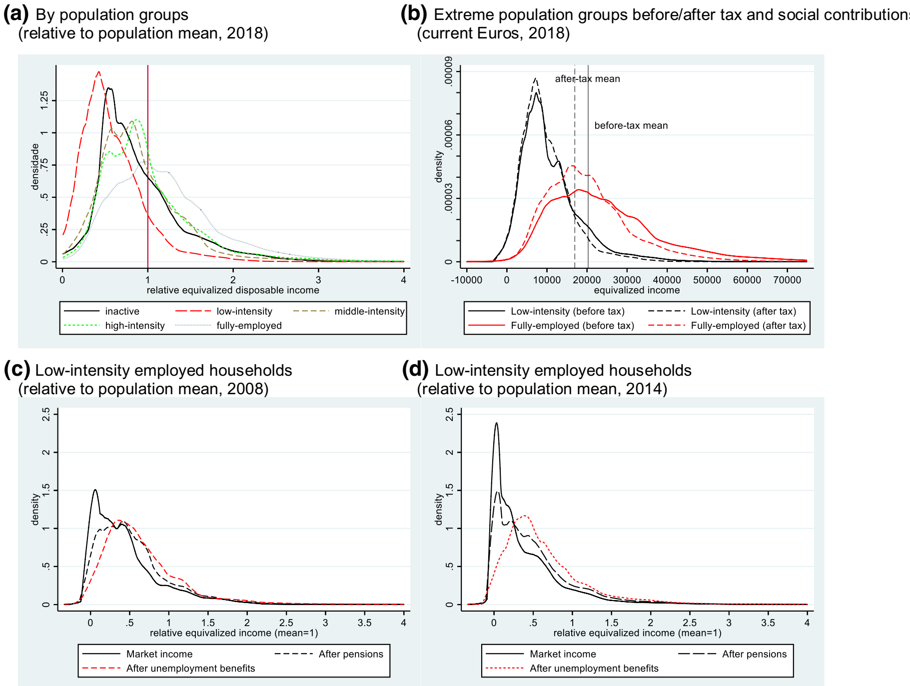

Figure 5. Densities for Disposable Income Distribution

Source: Author's construction based on microdata from ECV (INE). [Colour figure can be viewed at wileyonlinelibrary.com]

increase implies that they are moving away from the mean (either to the bottom in the case of relatively poor groups, or to the top in the case of richer groups). These changes could be the result of changes in the incomes of those initially in the group or of the different income levels of those entering or exiting the group.

There were, indeed, important distributional changes in the periods analysed (detailed densities for population groups by income source are displayed in the Appendix, Figures A1 and A2). Figure 6 shows an increase in the per capita contribution of households with low-intensity employment between 2011 and 2018 (0.022), along with an initial increase for the fully employed between 2011 and 2014 (0.024), followed by a reduction between 2014 and 2018 (0.019). However, the largest change is the decline in the average contribution of inactive households between 2008 and 2014 (0.076).

The role of income sources for the contributions of the different groups is illustrated in Figure 5b, which displays the densities before and after taxes and social security contributions for the two extreme groups in 2018: households with low-intensity employment and fully employed households. It is clear why the contribution of this source is higher among fully employed households as their incomes move closer to the middle of the distribution to a larger extent. That is, the contribution of the income source of a group just reflects its incidence in terms of how adding that source changes previous relative incomes and therefore its impact on inequality.

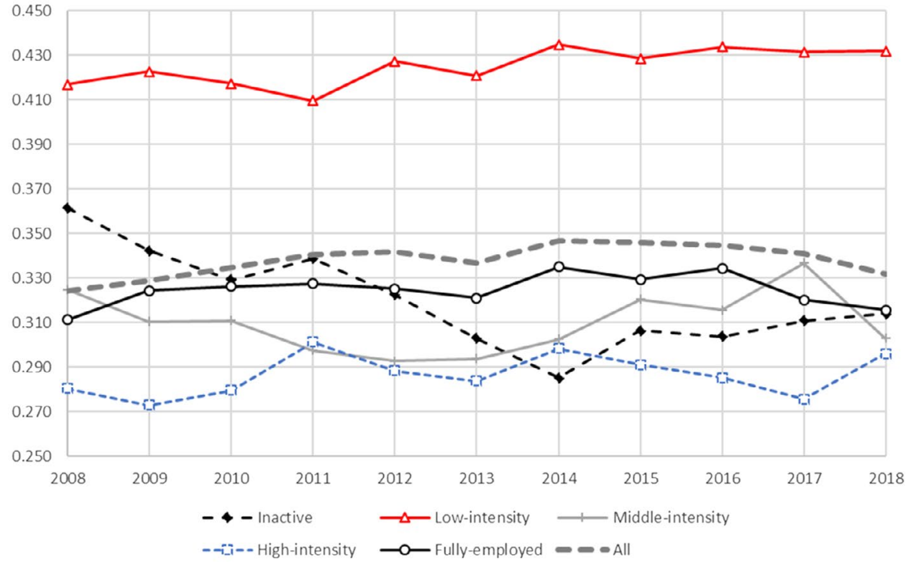

Figure 6. Per Capita Gini Group Contributions to Inequality in Disposable Income (β_k)
Source: Author's construction based on microdata from ECV (INE). [Colour figure can be viewed at wileyonlinelibrary.com]

The decomposition of the total change into compositional and distributional effects requires estimating the corresponding contribution matrices for the counterfactual distributions in which either the initial or the final population shares in each sub-period are kept constant. Tables 3 and 4 summarize the results of the decomposition by income source and by population group in each sub-period, while Tables A3 and A4 in the Appendix provide the more detailed decomposition by each pair of group and source. For robustness, the decomposition was undertaken with and without reweighting. The specification errors (due to nonlinearity in the relationship between inequality and group size) were found to be small and statistically insignificant in both sub-periods. For simplicity, we discuss here the case without reweighting.

5.2. *2008–14*

Most of the increase in inequality between 2008 and 2014 can be attributed to a compositional effect: 0.019–0.022 out of a 0.023 Gini increase, that is, 84 and 97 percent, depending on the counterfactual used as a reference. It is interesting to look at this decomposition in more detail, by income source and population group, as a small aggregate distributional effect can conceal large effects that offset each other, and that is indeed the case.

This large compositional effect reflects the net impact of the decline in the population share living in fully and highly employed households and the corresponding increase in the share of those with low-intensity employment, as the detailed decomposition confirms. For example, the compositional effect explains 0.076–0.079 out of 0.081 (i.e. 94–98 percent) of the increase in the contribution

TABLE 4

Blinder–Oaxaca RIF Decomposition of Changes in Gini (and Contributions by Groups and by Sources): Total, Compositional (CE), and Distributional (DE) Effects, 2008–14

	Counterfactual Distribution							
	Initial Group Shares, $\sum_k \frac{n_k}{n} \beta'_k$				Final Group Shares, $\sum_k \frac{n'_k}{n'} \beta_k$			
	CE	DE	Total	%CE/ Total	CE	DE	Total	%CE/ Total
Aggregate	0.019	0.004	0.023	84%	0.022	0.001	0.023	97%
$(I(y') - I(y))$	0.002	0.005	0.005		0.003	0.005	0.005	
By group $(S'_k - S_k)$								
Inactive	0.004	−0.010	−0.007	−58%	0.005	−0.011	−0.007	−73%
	0.001	0.001	0.002		0.002	0.001	0.002	
Low intensity	0.079	0.002	0.081	98%	0.076	0.005	0.081	94%
	0.003	0.001	0.004		0.004	0.004	0.004	
Middle intensity	0.005	−0.002	0.003	190%	0.005	−0.003	0.003	204%
	0.002	0.002	0.003		0.002	0.002	0.003	
High intensity	−0.017	0.003	−0.014	122%	−0.016	0.002	−0.014	115%
	0.002	0.002	0.003		0.002	0.002	0.003	
Fully employed	−0.053	0.012	−0.041	129%	−0.049	0.008	−0.041	120%
	0.003	0.004	0.004		0.003	0.002	0.004	
By source $(S'^j - S^j)$								
Market income	0.058	0.018	0.076	76%	0.051	0.025	0.076	67%
	0.003	0.005	0.006		0.004	0.006	0.006	
Pensions	−0.016	−0.009	−0.025	65%	−0.015	−0.010	−0.025	60%
	0.002	0.002	0.003		0.002	0.003	0.003	
Unemployment	−0.016	−0.004	−0.020	80%	−0.008	−0.012	−0.020	41%
	0.001	0.001	0.002		0.001	0.002	0.002	
Other social benefits	−0.006	0.001	−0.004	135%	−0.005	0.001	−0.004	115%
	0.001	0.001	0.001		0.001	0.002	0.001	
Taxes	−0.001	−0.003	−0.004	19%	−0.001	−0.003	−0.004	21%
	0.000	0.001	0.001		0.000	0.001	0.001	

Note: Detailed decomposition by each group and source, reported in Table A3 in the Appendix. The left and right panels correspond to the two decompositions in Section 3 using as counterfactual inequality $\sum_k \frac{n_k}{n} \beta'_k$ and $\sum_k \frac{n'_k}{n'} \beta_k$, respectively. Bootstraps standard errors below (1,000 replications).

Source: Author's construction based on microdata from ECV (INE).

of households with low-intensity employment, as well as the entire decline in the contributions of highly and fully employed households (compositional effect is larger than 100 percent). However, the equalizing contribution of inactive households over the period is not related to a change in the size of the group (that would explain a small increase in inequality) but with its smaller per capita contribution, that is, the group moved closer to the (shrinking) population mean.

Table 4 shows, by income source, that the large increase in market income inequality is also, to a large extent (0.051–0.058, i.e. 67–76 percent), a compositional effect. But this leaves a substantial distributional effect: between a quarter and a third of the higher disequalizing effect of market income is driven by changes in the market income distribution of groups and not in their size. Earnings inequality has been shown to run parallel to the evolution of the unemployment rate using tax files (2004–2010) and social security records (1988–2010) according to Bonhomme and Hospido (2013, 2017).

Table 4 also shows large compositional components explaining the more equalizing effect of adding other social benefits (more than 100 percent), pensions (60–65 percent) and unemployment benefits (40–80 percent). In contrast, only one-fifth of the larger equalizing effect of taxes during this period is the result of a compositional effect, the rest likely being the result of changes in the relative incomes of the employed households and of policy changes in the personal income tax to meet the fiscal balance challenges (with initial tax rises, reversed after 2015).

Among the distributional effects that can be identified during this period, the higher per capita contributions of fully employed households and market income stand out (their relative market incomes moved away from the population mean), in contrast to the smaller disequalizing effect of inactive households and the higher equalizing effect of pensions, unemployment benefits, and taxes (Table 4).

The more detailed decomposition by groups and sources (Table A3) reveals that the higher per capita contribution to the disposable income inequality of households either fully employed or with low-intensity employment was in fact, and not surprisingly, a contribution to higher market income inequality. The detailed analysis also identifies the income source of the smaller (i.e. negative sign) contributions of inactive households (pensions), households with low-intensity employment (unemployment benefits), and active households in general (taxes). For example, Figure 5c,d show that both pensions and unemployment benefits contribute to push the incomes of people living in low-intensity employed households towards the overall mean. While the impact of pensions is rather stable over time, the impact of unemployment benefits becomes more effective in 2014 when more people have access to this source of income. This is reflected in Table A3. In the case of old age pensions, the whole equalizing effect for this population group is composition effect (the group becomes larger), with no associated distributional effect. In the case of unemployment benefits, however, along the large composition effect, there is also an equalizing distributional effect in this period, especially if evaluated with 2014 population shares (with a higher unemployment rate).

5.3. *2014–2018*

A large share of the decline in inequality after 2014 (0.010–0.012 out of 0.015, 70–78 percent) was also driven by the compositional changes among active households during the recovery period, in which employment rates significantly improved, partially reversing the changes during the recession (Tables 4 and A4). The reduction in inequality during this period was indeed mostly driven by lower market income inequality, and this was entirely the result of a compositional effect. The decline in inequality was also associated with distributional changes, particularly a higher equalizing effect of pensions and other social benefits. At the most detailed level, there was a smaller per capita contribution of fully employed households (channeled through market income, pensions, and taxes), although with low statistical significance.

Some changes went in the opposite direction (increasing inequality) though, thus mitigating the reduction finally observed. About 36–61 percent of the contribution of unemployment benefits to higher inequality (0.005–0.009 out of 0.015) was also due to changes in the composition by groups in a context of declining

TABLE 5

Blinder–Oaxaca RIF Decomposition of Changes in Gini (and Contributions by Groups and by Sources): Total, Compositional (CE), and Distributional (SE) Effects, 2014–2018

	Counterfactual Distribution							
	Initial Group Shares, $\sum_k \frac{n_k}{n} \beta'_k$				Final Group Shares, $\sum_k \frac{n'_k}{n'} \beta_k$			
	CE	DE	Total	%CE/ Total	CE	DE	Total	%CE/ Total
Aggregate	−0.012	−0.003	−0.015	78%	−0.010	−0.004	−0.015	70%
$(I(y') - I(y))$	0.001	0.005	0.005		0.001	0.005	0.005	
By group $(S'_k - S_k)$								
Inactive	0.004	0.004	0.008	47%	0.003	0.005	0.008	43%
	0.002	0.001	0.002		0.001	0.001	0.002	
Low intensity	−0.044	−0.001	−0.045	98%	−0.044	−0.001	−0.045	99%
	0.004	0.003	0.004		0.004	0.002	0.004	
Middle intensity	0.000	0.000	0.000	111%	0.000	0.000	0.000	111%
	0.002	0.001	0.002		0.002	0.001	0.002	
High intensity	−0.003	0.000	−0.003	92%	−0.003	0.000	−0.003	93%
	0.002	0.002	0.002		0.002	0.002	0.002	
Fully employed	0.032	−0.006	0.026	125%	0.034	−0.008	0.026	133%
	0.003	0.003	0.004		0.003	0.003	0.004	
By source $(S'^j - S^j)$								
Market income	−0.021	0.003	−0.019	114%	−0.021	0.003	−0.019	114%
	0.003	0.005	0.006		0.003	0.005	0.006	
Pensions	0.001	−0.008	−0.007	−14%	0.000	−0.007	−0.007	5%
	0.002	0.003	0.003		0.002	0.002	0.003	
Unemployment	0.005	0.009	0.015	36%	0.009	0.006	0.015	61%
	0.001	0.002	0.002		0.001	0.001	0.002	
Other social	0.004	−0.005	−0.001	−338%	0.002	−0.004	−0.001	−232%
benefits	0.001	0.002	0.001		0.000	0.001	0.001	
Taxes	0.000	−0.002	−0.002	7%	0.000	−0.002	−0.002	−2%
	0.000	0.001	0.001		0.000	0.001	0.001	

Note: Detailed decomposition by each group and source, reported in Table A4 in the Appendix. The left and right panels correspond to the two decomposition in Section 3 using as counterfactual inequality $\sum_k \frac{n_k}{n} \beta'_k$ and $\sum_k \frac{n'_k}{n'} \beta_k$, respectively. Bootstraps standard errors below (1,000 replications).

Source: Author's construction based on microdata from ECV (INE).

unemployment rates. But there were also substantial changes in the distribution of the benefits received by households for other reasons because contributory unemployment benefits are in general subject to a limited duration (maximum 2 years), while non-contributory unemployment benefits that might replace them are less generous and subject to additional age and family requirements. Another small contribution towards higher inequality came from inactive households (through market income). With recovery, this group moved away again from the population mean.

It is interesting to note that the nearly zero distributional effect of households with low-intensity employment during this period (Table 5) is actually the result of two opposite forces that cancel each other out: a smaller equalizing effect of unemployment benefits and a higher equalizing effect of pensions and other benefits (Table A4).

6. Concluding Remarks

I have introduced a new approach for the simultaneous analysis of inequality by population groups and by income sources. In so doing, I have adopted the RIF approach to measure the contribution of a group to overall inequality and a sequential marginal approach to measure the corresponding contribution of an income source. Both approaches can be combined so that the contribution of a group through an income source is defined to be the change in the group contribution when that source is added. This approach thus provides a detailed decomposition of overall inequality into the contribution of each group and source and can be used to explain changes in inequality over time. The contribution of a group and income source will basically depend on the incidence of the income source among population groups, that is, how relative incomes change after adding the source. Instead of showing incidence curves, the method enables us to directly measure the impact on any inequality index and is thus a good complement to incidence analysis. The method is consistent with partial analysis of groups and/or sources, as well as with the use of RIF regressions to extend the Blinder–Oaxaca decomposition to a change in inequality into its compositional and structural components. While the former accounts for changes in inequality driven by changes in the composition of households by specific characteristics, the latter provides the impact of the change in their conditional income distribution.

The approach has been used to explain the trend in inequality in Spain after the outbreak of the Great Recession, identifying a large compositional effect at the aggregate level, although the detailed decomposition also enabled the unravelling of important distributional effects going in opposite directions, as well as factors that contributed to smooth the inequality trend.

Indeed, the increase in inequality during the recession (2008–2014) can be strongly associated with the shift in population from highly/fully employed households towards households with low-intensity employment, triggering an increase in market income inequality that was only partially mitigated by a higher equalizing effect of the other sources, especially old-age and survivor pensions and unemployment benefits. The higher market income inequality, as well as the higher equalizing effect of social benefits and taxes, was not only the result of the compositional effect, but also of distributional changes taking effect (in general, reinforcing the impact of the compositional effects). Inequality was later reduced (2014–18) due to the reversed process, with a shift in the population at that time from low-employed to fully employed households, triggering a decline in market income inequality, mitigated by lower equalizing effects of unemployment and other social benefits. The contribution of unemployment benefits to mitigate the decline in inequality was not just a compositional effect, but also the result of changes in distribution due to generally lower benefits.

It is interesting to note that while market income and unemployment benefits followed opposite trends along the economic cycle, the equalizing effect of pensions was more structural and tended to continuously increase over time, both during the recession and the recovery periods. Also, although the largest impact on inequality of adding pensions to household income was through inactive households, the effect of this source was partially channeled through group with the low-intensity employment during the recession and the fully employed group during the recovery, something which is less obvious. Other social benefits contributed to curb inequality in the recession, but the effect did not reverse during the recovery period, unlike what is reported for unemployment benefits. Taxes, however, contributed to reduce inequality in both periods, first through the low-intensity group and later through fully employed households.

References

Amuedo-Dorantes, C. and C. Borra, “Emerging Wealth Disparities After the Storm: Evidence from Spain,” *Review of Economics of the Household*, 109, 1119–49, 2018.

Blinder, A. S., “Wage Discrimination: Reduced Form and Distributional Estimates,” *Journal of Human Resources*, 8, 436–55, 1973.

Bonhomme, S. and L. Hospido, “Earnings Inequality in Spain: New Evidence Using Tax Data,” *Applied Economics*, 45, 4212–25, 2013.

______, “The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data,” *Economic Journal*, 127, 1244–78, 2017.

Chakravarty, S. R., *Inequality, Polarization and Poverty: Advances in Distributional Analysis*. Springer, New York, NY, 2009.

Chantreuil, F. and A. Trannoy, “Inequality Decomposition Values: The Trade-off Between Marginality and Efficiency,” *Journal of Economic Inequality*, 11, 83–98, 2013.

Charpentier, A. and S. Mussard, “Income Inequality Games,” *Journal of Economic Inequality*, 9, 529–54, 2011.

Cowell, F. A., and M. P. Victoria-Feser, “Robustness Properties of Inequality Measures: The Influence Function and the Principle of Transfers,” *Econometrica*, 64, 77–101, 1996.

DiNardo, J., N. M. Fortin, and T. Lemieux, “Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach,” *Econometrica*, 64, 1001–44, 1996.

Diris, R., F. Vandenbroucke, and G. Verbist, “The Impact of Pensions, Transfers and Taxes on Child Poverty in Europe: The Role of Size, Pro-Poorness and Child Orientation,” *Socio-Economic Review*, 15, 745–75, 2017.

Essama-Nssah, B. and P. J. Lambert, “Influence Functions for Policy Impact Analysis,” in J. A. Bishop and R. Salas (eds), *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber*, Chapter 6 (Research on Economic Inequality, 20). Emerald, Bingley, 135–59, 2012.

Eurostat, *Eurostat Database*, https://ec.europa.eu/eurostat (accessed August 2019), n.d.

Eurostat, *Methodological Guidelines and Description of EU-SILC Target Variables*, DocSILC065 (2017 Operation), European Commission Eurostat, Directorate F: Social Statistics, Unit F- 4: Quality of Life, Brussels, 2017.

Fields, G. S., “Accounting for Income Inequality and its Change: A New Method with Application to U.S. Earnings Inequality,” in S. W. Polacheck (ed.), *Worker Well-Being and Public Policy* (Research in Labor Economics), Vol. 22, Emerald, Bingley, 1–38, 2003.

Firpo, S., N. M. Fortin, and T. Lemieux, *Decomposing Wage Distributions Using Recentered Influence Function Regressions*, Unpublished Manuscript, Vancouver, University of British Columbia, 2007.

______, “Unconditional Quantile Regressions,” *Econometrica*, 77, 953–73, 2009.

Fortin, N. M., T. Lemieux, and S. Firpo, “Decomposition Methods in Economics,” in O. Ashenfelter and D. Card, (eds), *Handbook of Labor Economics*, North Holland, Amsterdam, 1–102, 2011.

Gradín, C., “Quantifying the Contribution of a Subpopulation to Inequality: An Application to Mozambique,” *Journal of Economic Inequality*, forthcoming, 2020.

______, “Why is Income Inequality So High in Spain?,” in L. Cappellari, S. Polachek, and K. Tatsiramos (eds), *Inequality Around the World. Research on Labor Economics*, Emerald, Bingley, 44, 109–77, 2016.

Gradín, C., O. Cantó, and C. Del Río, “Unemployment and Spell Duration During the Great Recession in the EU,” *International Journal of Manpower*, 36, 216–35, 2015.

______, “Measuring Employment Deprivation in the EU Using a Household-level Index,” *Review of Economics of the Household*, 15, 639–67, 2017.

Hampel, F. R., “The Influence Curve and Its Role in Robust Estimation,” *Journal of the American Statistical Association*, 60, 383–93, 1974.

INE, *INEBase*, https://www.ine.es/dyngs/INEbase/listaoperaciones.htm (accessed September 2019), n.d.

______, *Encuesta de Condiciones de Vida: Metodología*, Instituto Nacional de Estadística, Madrid, 2019. https://www.ine.es/daco/daco42/condivi/ecv_metodo.pdf.
Kakwani, N. C., "Measurement of Tax Progressivity: An International Comparison," *Economic Journal*, 87, 71–8, 1977.
Monti, A. C., "The Study of the Gini Concentration Ratio by Means of the Influence Function," *Statistica*, 51, 561–77, 1991.
Morduch, J., and T. Sicular, "Rethinking Inequality Decomposition, with Evidence from Rural China," *Economic Journal*, 112, 93–106, 2002.
Musgrave, R. A., and T. Thin, "Income Tax Progression 1929–48," *Journal of Political Economy*, 56, 498–514, 1948.
Oaxaca, R. L., "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693–709, 1973.
Oaxaca, R. L., and M. R. Ransom, "Identification in Detailed Wage Decompositions," *Review of Economics and Statistics*, 81, 154–7, 1999.
Sastre, M., and A. Trannoy, "Shapley Inequality Decomposition by Factor Components: Some Methodological Issues," *Journal of Economics*, 77, 51–8, 2002.
Shorrocks, A. F., "Inequality Decomposition by Factor Components," *Econometrica*, 50, 193–211, 1982.
______, "Inequality Decomposition by Population Subgroups," *Econometrica*, 52, 1369–85, 1984.
______, "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," *Journal of Economic Inequality*, 11, 99–126, 2013.
Verbist, G., R. Diris, and F. Vandenbroucke, "Solidarity Between Generations in Extended Families. Direction, Size and Intensity," University of Antwerp Herman Deleeck Centre for Social Policy, Working Paper N18.16, 2018.
Victoria-Feser, M. P., "Robust Methods for Personal Income Distribution Models," Ph.D. Thesis, University of Geneva, Thesis No. 384, 1993.
Victoria-Feser, M. P. and E. Roncherri, "Robust Methods for Personal Income Distribution Models," *Canadian Journal of Statistics*, 22, 247–58, 1994.
Yun, M.-S., "Earnings Inequality in USA, 1969–99: Comparing Inequality Using Earnings Equations," *Review of Income and Wealth*, 52, 127–44, 2006.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Table A.1a: Mean Equivalized Disposable Income, Total and by Sources

Table A.1b: Mean Equivalized Disposable Income, Total and by Sources and Population Groups

Table A.1c: Shapley Inequality Contribution of Social Benefits to Changes in Inequality

Table A.2: Gini Contributions by Group and Income Source (S_k^j)

Table A.3: Detailed Blinder–Oaxaca Decomposition of Inequality Change ($S'^j_k - S^j_k$), 2008–2014

Table A.4: Detailed Blinder–Oaxaca Decomposition of Inequality Change ($S'^j_k - S^j_k$), 2014–2018

Figure A.1: Densities of Relative Disposable Income (Mean = 1) Over Time: by Population Group Inactive

Figure A.2: Population Group Densities of Relative Disposable Income (Mean = 1): Different Years 2008