Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

What Lies Between Groups? A Distribution-Sensitive Gini Decomposition

Vito De Sandi

vito.desandi@liser.lu

Luxembourg Institute of Socio-Economic Research

Research Article

Keywords: Gini coefficient, inequality decomposition, subgroup inequality, distributional overlap, regional inequality

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11259970/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

What Lies Between Groups? A Distribution-Sensitive Gini Decomposition

Vito De Sandi*

Luxembourg Institute of Socio Economic Research

Abstract

Conventional inequality subgroup decompositions often summarize between-group inequality through differences in group means. This is informative about differences in expected outcomes, but not about differences between the underlying group-specific distributions. The paper introduces a decomposition of the Gini index to address this concern. For each pair of groups, I identify the maximal subdensity shared by their outcome distributions and decompose the cross-group Gini interaction into a common-mass component and a distributional-separation component. The resulting decomposition is non-negative and preserves the conventional within-group contribution. An application to regional income inequality in five European countries shows that distribution-specific mass can account for a quantitatively relevant part of the national Gini coefficient even when conventional between-region components are small.

Keywords: Gini coefficient; inequality decomposition; subgroup inequality; distributional overlap; regional inequality.
JEL classification: D31; D63; C43; R12.

1 Introduction and motivation

A large literature studies how aggregate inequality is distributed within and across population subgroups (Bourguignon, 1979; Cowell, 1980; Shorrocks, 1980, 1984). For additively decomposable inequality measures, the standard approach separates a within-group contribution from a between-group component associated with differences in subgroup mean outcomes (Shorrocks, 1980). The Gini coefficient poses a more difficult subgroup-decomposition problem because cross-group comparisons contain information that is not exhausted by differences in subgroup means. A convenient Gini decomposition is

$$G = G^W + G^B + G^R, \tag{1}$$

where G^W is the conventional same-group contribution, G^B captures differences in subgroup means, and G^R collects the remaining cross-group interaction (Pyatt, 1976; Silber, 1989; Yitzhaki and Lerman, 1991; Lambert and Aronson, 1993). Several contributions, including Dagum (1997), has emphasized that this interaction depends on the relative position and overlap of subgroup distributions.

Recent contributions make the distributional limitation of mean-based comparisons particularly explicit. Attili (2024) shows that groups with the same mean can differ substantially in dispersion, skewness, support, or other features of their distributions, and proposes a Gini decomposition that

*vito.desandi@liser.lu

compares the full subgroup distributions. Heikkuri and Schief (2026) instead derive an axiomatic two-component Gini decomposition in which between-group inequality is sensitive to the full distributions and vanishes when those distributions coincide. The contribution here addresses a different aspect of the problem: the internal composition of the cross-group Gini interaction itself.

The key point is that a positive cross-group Gini interaction does not necessarily imply that the underlying group distributions differ. If Y_k and Y_ℓ are independently drawn from two identical non-degenerate distributions, then $E|Y_k - Y_\ell|$ will generally be positive even though group membership contains no distributional information. Part of the cross-group interaction can therefore reflect ordinary interpersonal dispersion generated by distributional mass shared by the groups. Another part can arise because at least one observation is drawn from probability mass that is specific to one group distribution.

The proposed decomposition separates these two sources. For every pair of subgroup densities, it identifies their maximal common subdensity and the remaining group-specific mass. The cross-group Gini interaction is then decomposed into a common-mass component (G^C), generated when both observations are drawn from mass present in both groups, and a distributional-separation component (G^S), generated by comparisons in which at least one observation is drawn from group-specific mass. The use of the maximal common subdensity identifies the largest part of the two distributions that can be regarded as shared by both groups, leaving the remaining mass necessarily group-specific.

A Gini index decomposed as proposed here does not replace the conventional between-mean component, nor does it reinterpret every cross-group difference as evidence of group separation. Rather, it reorganizes the cross-group interaction according to a different principle: it distinguishes dispersion generated by common distributional mass from dispersion involving group-specific mass.

This distinction is potentially relevant whenever populations are partitioned into economically meaningful groups and the object of interest extends beyond differences in average outcomes. Spatial inequality provides a particularly useful application. A recurring result in the empirical literature is that the large majority of national income inequality is attributed to the within-region component, whereas conventional between-region inequality is comparatively small (Shorrocks and Wan, 2005; Novotný, 2007; International Monetary Fund, 2019; OECD, 2026) (see Appendix B for an overview of these results). This finding is often interpreted as evidence that the spatial dimension of inequality is of limited quantitative importance.

However, the familiar result that inequality is "mostly within regions" should be interpreted with care. Its predominance may reflect genuine heterogeneity among individuals within regions, but it also reflects the construction of the decomposition itself: distributional differences that do not manifest themselves through regional means cannot enter the conventional between component.

At the same time, replacing regional means with the complete cross-region Gini interaction does not solve the measurement problem. Let

$$\Delta_{rs} = E|Y_r - Y_s|$$

denote the cross-region absolute-distance interaction. As discussed above, this interaction can remain positive even when the two regional distributions coincide. A large cross-region interaction therefore need not imply strong regional separation, just as a small mean-based between-region component need not imply similar regional distributions.

The proposed decomposition distinguishes these cases by separating the part accounted for by

common distributional mass, G^C, from the part involving region-specific mass, G^S.

We illustrate the empirical relevance of this distinction using harmonized microdata from the Luxembourg Income Study (LIS). The application covers five European countries over a ten-year period and shows how the proposed decomposition changes the interpretation of spatial inequality relative to the conventional mean-based approach. In particular, it allows us to assess whether observed cross-region inequality primarily reflects differences between region-specific parts of the distributions or interactions involving income mass that is common across regions.

The rest of the paper proceeds as follows. Section 2 develops the Gini-compatible common-mass and distributional-separation decomposition. Section 3 applies the decomposition to five European countries and compares its results with those obtained from the standard approach. Section 4 concludes.

2 A distributional decomposition of the Gini

2.1 Preliminaries

Let N be a population partitioned into K mutually exclusive groups, indexed by $k = 1, \ldots, K$. Group k has population share w_k, with $\sum_{k=1}^{K} w_k = 1$. Let $Y_k \sim F_k$ denote the non-negative outcome of a randomly drawn individual from group k, where F_k is absolutely continuous with density f_k. The mean outcome in group k is

$$\mu_k = \mathbb{E}[Y_k] = \int_0^\infty y f_k(y)\, dy.$$

The aggregate mean is

$$\mu = \sum_{k=1}^{K} w_k \mu_k.$$

Let $Y_k \sim F_k$ denote the outcome of a randomly drawn individual from group k. The overall Gini coefficient can be written as

$$G = \frac{1}{2\mu} \sum_{k=1}^{K} \sum_{\ell=1}^{K} w_k w_\ell E|Y_k - Y_\ell|. \tag{2}$$

Separating the within-group and cross-group terms gives

$$G = \sum_{k=1}^{K} w_k^2 \frac{\mu_k}{\mu} G_k + \frac{1}{\mu} \sum_{k<\ell} w_k w_\ell \Delta_{k\ell}, \tag{3}$$

where

$$G_k = \frac{1}{2\mu_k} \int_0^\infty \int_0^\infty |x - y| f_k(x) f_k(y)\, dx\, dy \tag{4}$$

is the Gini coefficient within group k, and

$$\Delta_{k\ell} = E|Y_k - Y_\ell| = \int_0^\infty \int_0^\infty |x - y| f_k(x) f_\ell(y)\, dx\, dy \tag{5}$$

is the cross-group Gini interaction. Define

$$G^W = \sum_{k=1}^{K} w_k^2 \frac{\mu_k}{\mu} G_k. \tag{6}$$

The conventional between-group mean component is

$$G^B = \frac{1}{\mu} \sum_{k<\ell} w_k w_\ell |\mu_k - \mu_\ell|, \tag{7}$$

so that the standard representation is

$$G = G^W + G^B + G^R, \tag{8}$$

with

$$G^R = \frac{1}{\mu} \sum_{k<\ell} w_k w_\ell \left[\Delta_{k\ell} - |\mu_k - \mu_\ell|\right]. \tag{9}$$

2.2 Decomposition

We now decompose $\Delta_{k\ell}$ according to whether the relevant distributional mass is common to the two groups or specific to one of them. For each pair k, ℓ, define

$$m_{k\ell}(y) = \min\{f_k(y), f_\ell(y)\},$$

$$a_{k\ell}(y) = [f_k(y) - f_\ell(y)]_+, \qquad b_{k\ell}(y) = [f_\ell(y) - f_k(y)]_+,$$

where $[z]_+ = \max\{z, 0\}$. Hence

$$f_k(y) = m_{k\ell}(y) + a_{k\ell}(y), \qquad f_\ell(y) = m_{k\ell}(y) + b_{k\ell}(y). \tag{10}$$

The overlap and non-overlap masses are

$$\mathrm{OV}_{k\ell} = \int_0^\infty m_{k\ell}(y)\, dy, \qquad d_{k\ell} = 1 - \mathrm{OV}_{k\ell}, \tag{11}$$

and equivalently

$$d_{k\ell} = \int_0^\infty a_{k\ell}(y)\, dy = \int_0^\infty b_{k\ell}(y)\, dy = \frac{1}{2} \int_0^\infty |f_k(y) - f_\ell(y)|\, dy. \tag{12}$$

The quantity $OV_{k\ell}$ is the coefficient of overlapping introduced by Weitzman (1970), which measures the common area under two probability densities; see also Schmid and Schmidt (2006). For the interior case $0 < d_{k\ell} < 1$, define the normalized common density

$$h_{k\ell}(y) = \frac{m_{k\ell}(y)}{1 - d_{k\ell}},$$

and the normalized non-common densities

$$p_{k\ell}(y) = \frac{a_{k\ell}(y)}{d_{k\ell}}, \qquad q_{k\ell}(y) = \frac{b_{k\ell}(y)}{d_{k\ell}}.$$

Then

$$f_k(y) = (1 - d_{k\ell})h_{k\ell}(y) + d_{k\ell}p_{k\ell}(y),$$

$$f_\ell(y) = (1 - d_{k\ell})h_{k\ell}(y) + d_{k\ell}q_{k\ell}(y).$$

Substitution into the cross-group interaction yields

$$\begin{aligned}\Delta_{k\ell} = (1 - d_{k\ell})^2 A_{k\ell}^{HH} + d_{k\ell}(1 - d_{k\ell})A_{k\ell}^{HQ} \\ + d_{k\ell}(1 - d_{k\ell})A_{k\ell}^{PH} + d_{k\ell}^2 A_{k\ell}^{PQ},\end{aligned} \tag{13}$$

where

$$\begin{aligned}A_{k\ell}^{HH} &= \int_0^\infty \int_0^\infty |x - y| h_{k\ell}(x)h_{k\ell}(y)\, dx\, dy, \\ A_{k\ell}^{HQ} &= \int_0^\infty \int_0^\infty |x - y| h_{k\ell}(x)q_{k\ell}(y)\, dx\, dy, \\ A_{k\ell}^{PH} &= \int_0^\infty \int_0^\infty |x - y| p_{k\ell}(x)h_{k\ell}(y)\, dx\, dy, \\ A_{k\ell}^{PQ} &= \int_0^\infty \int_0^\infty |x - y| p_{k\ell}(x)q_{k\ell}(y)\, dx\, dy.\end{aligned}$$

For $0 < d_{k\ell} < 1$, define

$$C_{k\ell} = (1 - d_{k\ell})^2 A_{k\ell}^{HH}, \tag{14}$$

and

$$S_{k\ell} = d_{k\ell}(1 - d_{k\ell}) \left(A_{k\ell}^{HQ} + A_{k\ell}^{PH}\right) + d_{k\ell}^2 A_{k\ell}^{PQ}. \tag{15}$$

Thus $\Delta_{k\ell} = C_{k\ell} + S_{k\ell}$. At the two boundaries, define

$$C_{k\ell} = \Delta_{k\ell}, \quad S_{k\ell} = 0 \qquad \text{if } d_{k\ell} = 0,$$

and

$$C_{k\ell} = 0, \quad S_{k\ell} = \Delta_{k\ell} \qquad \text{if } d_{k\ell} = 1.$$

Equivalently,

$$C_{k\ell} = \begin{cases} \Delta_{k\ell}, & d_{k\ell} = 0, \\ (1 - d_{k\ell})^2 A_{k\ell}^{HH}, & 0 < d_{k\ell} < 1, \\ 0, & d_{k\ell} = 1, \end{cases} \tag{16}$$

and

$$S_{k\ell} = \begin{cases} 0, & d_{k\ell} = 0, \\ d_{k\ell}(1 - d_{k\ell}) \left(A_{k\ell}^{HQ} + A_{k\ell}^{PH}\right) + d_{k\ell}^2 A_{k\ell}^{PQ}, & 0 < d_{k\ell} < 1, \\ \Delta_{k\ell}, & d_{k\ell} = 1. \end{cases} \tag{17}$$

By construction,

$$\Delta_{k\ell} = C_{k\ell} + S_{k\ell} \qquad \text{for every } d_{k\ell} \in [0, 1]. \tag{18}$$

The component $C_{k\ell}$ is the contribution to the cross-group absolute-distance interaction generated by common density mass, whereas $S_{k\ell}$ is the contribution involving density mass that is specific to at

least one group[1].

Define the aggregate components

$$G^C = \frac{1}{\mu}\sum_{k<\ell} w_k w_\ell C_{k\ell}, \qquad G^S = \frac{1}{\mu}\sum_{k<\ell} w_k w_\ell S_{k\ell}. \tag{19}$$

Theorem 1 (Gini common-mass and distributional-separation decomposition)**.** *Suppose that the population is partitioned into K mutually exclusive groups, each group-specific distribution is absolutely continuous on $[0,\infty)$, and each has a finite and strictly positive mean. Then*

$$G = G^W + G^C + G^S, \tag{20}$$

with $G^W \geq 0$, $G^C \geq 0$, and $G^S \geq 0$. The proof is given in Appendix A.

2.3 Interpretation and discussion

The distinction between cross-group dispersion, differences in group means, and differences in group distributions can be illustrated directly through the geometry of the group-specific densities. Figure 1 considers three cases.

For each pair of groups, the common subdensity is

$$m_{k\ell}(y) = \min\{f_k(y), f_\ell(y)\}, \tag{21}$$

while $d_{k\ell}$ measures the probability mass that is not shared by the two distributions. Thus, $d_{k\ell} = 0$ when the distributions are identical, whereas $d_{k\ell} > 0$ indicates that some probability mass is group-specific.

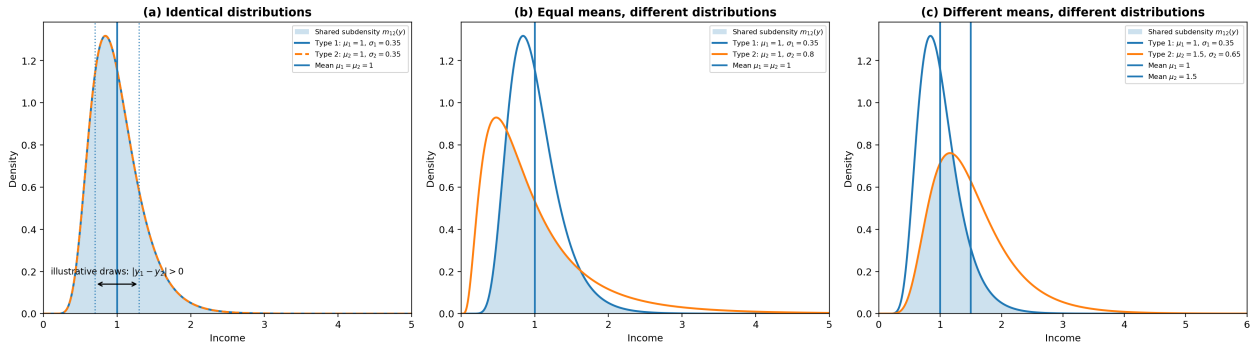


Figure 1: Cross-group dispersion and distributional separation. Panel (a) shows that the cross-group interaction can be positive even when the two group distributions are identical. Panel (b) shows that equal group means do not imply identical distributions. Panel (c) shows that when group means differ, the mean difference captures only one dimension of the difference between the two distributions. The shaded region represents the common subdensity $m_{12}(y) = \min\{f_1(y), f_2(y)\}$.

Identical distributions. Panel (a) provides the benchmark case. Suppose that

$$F_1 = F_2. \tag{22}$$

[1]Notice also that, under the maintained absolute-continuity assumption, $S_{k\ell} = 0$ if and only if $F_k = F_\ell$. Hence the separation component vanishes precisely when the two group-specific distributions coincide.

The two densities then coincide completely, so that all probability mass is common to the two groups. Hence,

$$d_{12} = 0. \tag{23}$$

Nevertheless, two individuals drawn independently from the same distribution need not have the same outcome. Therefore,

$$\Delta_{12} = E|Y_1 - Y_2| > 0 \tag{24}$$

whenever the common distribution has positive dispersion. The two vertical dotted lines in panel (a) illustrate two possible draws from the same distribution with different outcomes.

In this case,

$$S_{12} = 0, \qquad C_{12} = \Delta_{12}. \tag{25}$$

Thus, a positive cross-group interaction does not by itself imply that the two group distributions differ.

Equal means, different distributions. Panel (b) considers the case in which the two groups have the same mean but different distributions,

$$\mu_1 = \mu_2, \qquad F_1 \neq F_2. \tag{26}$$

The pairwise mean difference is therefore

$$|\mu_1 - \mu_2| = 0. \tag{27}$$

A comparison based only on group means would consequently indicate no difference between the two groups. The figure shows, however, that their densities do not coincide. Only part of the probability mass is shared, while the remaining mass is specific to one of the two groups. Hence,

$$d_{12} > 0, \tag{28}$$

and the distributional-separation component S_{12} may also be positive.

This case illustrates an important distinction: equality of expected outcomes does not imply equality of outcome distributions. Two groups can have the same mean while differing in dispersion, shape, support, or other features of their distributions. These differences are not captured by $|\mu_1 - \mu_2|$, but they can generate a positive S_{12}.

Different means and different distributions. Panel (c) considers the case in which both the group means and the distributions differ. Here,

$$|\mu_1 - \mu_2| > 0. \tag{29}$$

The conventional between-group component therefore captures a difference in average outcomes. However, the difference between μ_1 and μ_2 describes only one aspect of the difference between the two distributions. The densities may also differ in dispersion, shape, support, and the allocation of

probability mass across the outcome space.

Interpretation of the decomposition. More generally, for each pair of groups,

$$\Delta_{k\ell} = C_{k\ell} + S_{k\ell}. \tag{30}$$

The two components distinguish between two sources of cross-group dispersion. The component $C_{k\ell}$ is generated when both observations come from probability mass that is common to the two group distributions. In contrast, $S_{k\ell}$ collects all comparisons in which at least one observation comes from probability mass that is specific to one of the groups.

The proposed decomposition therefore complements the conventional subgroup decomposition by asking how much of the cross-group interaction is generated by probability mass shared by the groups and how much involves probability mass that is specific to at least one group.

At the aggregate level,

$$G^S = \frac{1}{\mu} \sum_{k<\ell} w_k w_\ell S_{k\ell} \tag{31}$$

summarizes the contribution associated with distribution-specific mass across all pairs of groups. A small value of G^S indicates that relatively little of the cross-group Gini interaction involves group-specific probability mass, whereas a larger value indicates greater distributional separation across groups.

Relation to the conventional between-group component. The distributional-separation component also nests the conventional between-group mean component in the following sense:

$$S_{k\ell} \geq |\mu_k - \mu_\ell|$$

for every pair of groups. Consequently,

$$G^S \geq G^B.$$

Defining

$$G^D \equiv G^S - G^B \geq 0,$$

the proposed and conventional decompositions admit the common refinement

$$G = G^W + G^B + G^C + G^D,$$

so that

$$G^R = G^C + G^D, \qquad G^S = G^B + G^D.$$

Thus, G^D measures the distribution-sensitive separation contribution in excess of that represented by differences in group means. (Proof in Appendix A).

3 Empirical application

3.1 Data

The main cross-sectional application uses Luxembourg Income Study (LIS) microdata for Belgium, Germany, France, Italy, and Luxembourg in 2020. To assess whether the 2020 results reflect a persistent feature of regional inequality, the same baseline specification is additionally estimated for 2010, 2012, 2014, 2016, and 2018. The outcome is non-negative annual disposable household income, equivalised using the square-root equivalence scale. No top- or bottom-coding is applied. Geographic groups are defined using the harmonized LIS variable `region_c`.[2]

For each country, national inequality is reported under three approaches. The proposed Gini decomposition separates G into G^W, G^C, and G^S. The conventional Gini representation reports G^W, the mean-based between-region component G^B, and the residual interaction G^R. The Theil index is decomposed into its standard within- and between-region components.

3.2 Results

Table 1 reports the 2020 decomposition in both relative and absolute terms. For each component, the first line gives its percentage share of the corresponding national inequality index, while the value in parentheses below gives the component contribution in index units. The columns G and T report the national Gini and Theil levels, respectively. The empirical implementation of the proposed decomposition requires estimation of the region-specific income densities. All estimates use the baseline empirical specification with $J = 1{,}000$ grid points, the `nrd0` bandwidth rule, and the direct weighted empirical estimator of the pairwise cross-region interaction $\Delta_{k\ell}$. The estimation procedure, including the weighted kernel density estimator, numerical integration, and construction of the pairwise common-mass and separation components, is described in Appendix D.

Table 1: Regional decomposition of income inequality, 2020

		Proposed Gini			Conventional Gini			Theil		
Country	G	G^W	G^C	G^S	G^W	G^B	G^R	T	Within	Between
		43.2	37.3	19.5	43.2	13.7	43.1		98.0	2.0
Belgium	0.262	(0.113)	(0.0979)	(0.051)	(0.113)	(0.0359)	(0.113)	0.127	(0.125)	(0.00251)
		11.7	66.5	21.8	11.7	14.6	73.8		98.0	2.0
Germany	0.320	(0.0373)	(0.213)	(0.0697)	(0.0373)	(0.0465)	(0.236)	0.202	(0.198)	(0.00398)
		8.4	63.5	28.1	8.4	18.2	73.4		96.9	3.1
France	0.297	(0.0248)	(0.188)	(0.0834)	(0.0248)	(0.054)	(0.218)	0.179	(0.173)	(0.00551)
		7.1	52.3	40.7	7.1	32.3	60.6		89.9	10.1
Italy	0.328	(0.0232)	(0.171)	(0.133)	(0.0232)	(0.106)	(0.199)	0.189	(0.17)	(0.0191)
		15.6	54.8	29.6	15.6	21.2	63.3		95.6	4.4
Luxembourg	0.281	(0.0438)	(0.154)	(0.0831)	(0.0438)	(0.0594)	(0.178)	0.135	(0.129)	(0.00593)

Notes: For each decomposition component, the first line reports its percentage share of the corresponding national inequality index; the value in parentheses reports the contribution in index units. G is the national Gini coefficient and T is the national Theil index. The raw proposed-Gini components satisfy $G = G^W + G^C + G^S$ up to numerical precision; the conventional Gini components satisfy $G = G^W + G^B + G^R$.

The index values in parentheses make clear that the percentage decomposition is not driven by a common national inequality level across countries. In 2020, the national Gini ranges from 0.262 in Belgium to 0.328 in Italy, while the proposed components in index units sum to these national totals.

[2]Luxembourg Income Study Database: `https://www.lisdatacenter.org/`.

The conventional mean-based between component G^B is comparatively limited in all five countries, ranging from 13.7 percent of total Gini inequality in Belgium to 32.3 percent in Italy. The standard Theil between-region component is smaller still: approximately 2 percent in Belgium and Germany, below 5 percent in France and Luxembourg, and 10.1 percent in Italy. These findings are consistent with the familiar conclusion from the spatial-inequality literature that regional mean differences account for a relatively small share of aggregate interpersonal inequality.

The proposed decomposition gives a different interpretation of the same cross-region Gini interaction. A substantial fraction is generated by common distributional mass: G^C accounts for 66.5 percent of total Gini inequality in Germany and 63.5 percent in France. This part should not be read as substantive distributional separation between regions, because it represents dispersion generated when both observations come from income mass shared by the relevant regional distributions.

At the same time, G^S is quantitatively relevant in every country, ranging from 19.5 percent in Belgium to 40.7 percent in Italy.

This component identifies cross-region dispersion involving probability mass that is not common to the relevant regional distributions and therefore captures information that regional means alone cannot recover. Regarding Italy, the conventional between-mean component accounts for 32.3 percent of total Gini inequality, while the distributional-separation component G^S accounts for 40.7 percent. Since $G^S \geq G^B$, these quantities should not be interpreted as unrelated measures of cross-region differences. Rather, the conventional mean component provides a lower-bound benchmark for distributional separation. The additional distribution-sensitive contribution beyond regional mean differences, $G^S - G^B$, is approximately 8.4 percentage points of the national Gini. Regional mean differences therefore account for a substantial part of measured regional separation, while a further quantitatively meaningful component reflects differences not summarized by regional averages.

Figure 2 gives a clearer comparison of the three approaches. The standard Theil decomposition attributes the overwhelming majority of inequality to the within-region component in every country. The conventional Gini decomposition produces a much larger cross-region interaction, but only a comparatively small part is captured by the mean-based component G^B; a large fraction remains in the residual G^R. The proposed decomposition reorganizes this same Gini interaction by separating the common-mass component G^C from the distributional-separation component G^S.

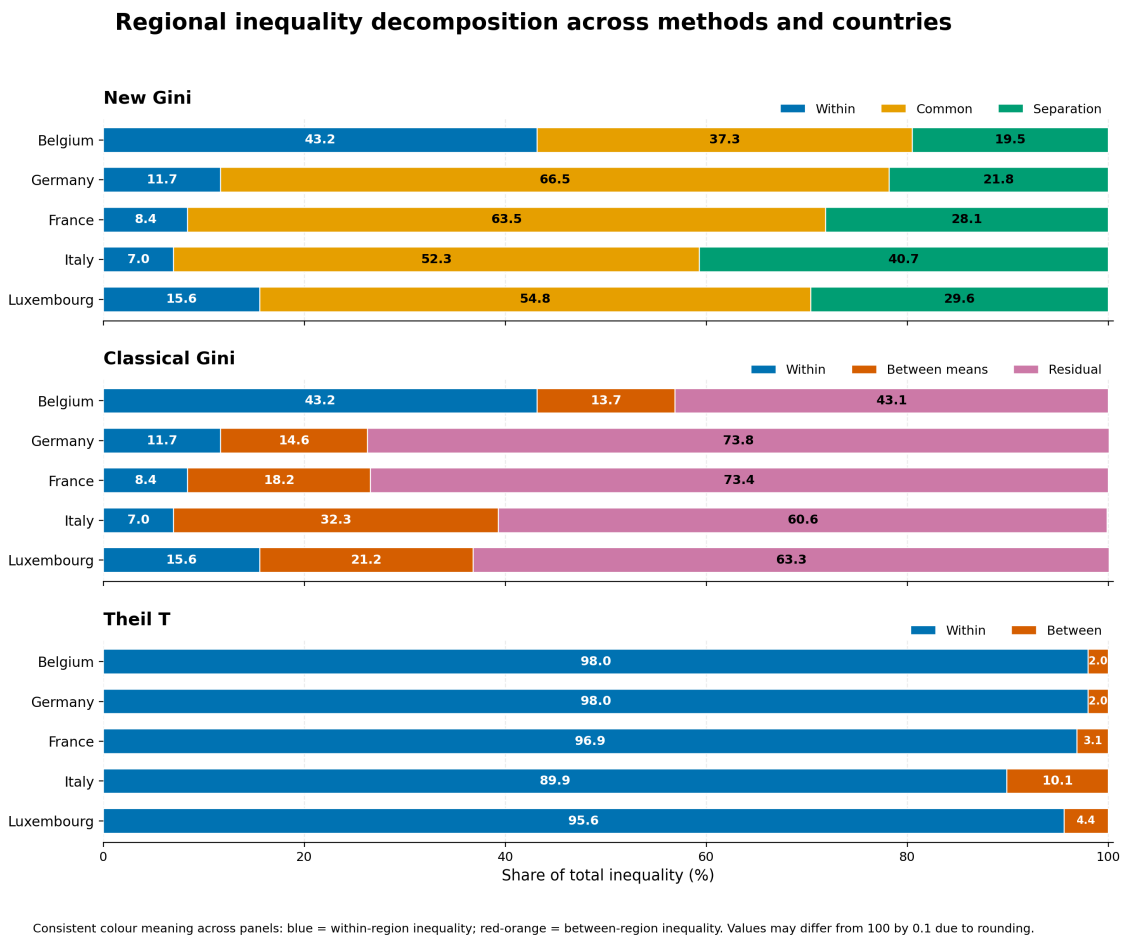


Figure 2: Regional inequality decomposition across methods and countries. Shares are percentages of the relevant national inequality index and each bar sums to 100.

The comparison also clarifies how the component shares should be read. In Germany and France, G^C is much larger than G^S, so a large part of the cross-region interaction is generated by income mass shared across regional distributions and should not be interpreted as strong regional separation. At the same time, G^S is positive and quantitatively relevant in every country, with the largest contribution in Italy. Regional averages may therefore be relatively close even when the distributions allocate non-negligible probability mass differently across the income range.

This is why the spatial measurement problem is economically relevant. A small conventional between-region component supports the statement that regional mean differences are limited; it does not support the stronger statement that region is unimportant for the income distribution. The reverse error is also possible: a large cross-region Gini interaction is not automatically a regional policy problem when much of it is G^C. The decomposition therefore does not establish causality or prescribe an intervention; it shows that the distributional relevance of geography cannot be assessed from regional means alone.

3.2.1 Evolution over time

Figure 3 extends the baseline decomposition to 2010, 2012, 2014, 2016, and 2018, while retaining 2020 as the benchmark year.

The temporal extension shows that the main cross-country patterns observed in 2020 are persistent. Belgium is characterized by a comparatively large within-region component throughout the period:

G^W remains between 42.0 and 43.2 percent of total Gini inequality, while G^C fluctuates between 37.3 and 39.8 percent and G^S between 17.8 and 19.6 percent. France and Germany display a substantially smaller within-region component, close to 8 percent and 12 percent respectively, together with a dominant common-mass component. In France, G^S rises from 25.3 percent in 2010 to 28.1 percent in 2020 and reaches 28.7 percent in 2018. In Germany, G^S varies between 20.4 and 25.7 percent over the observed years.

Italy continues to stand out for the importance of distribution-specific regional mass. Between 2010 and 2016, G^S accounts for between 42.3 and 44.7 percent of national Gini inequality, compared with approximately 48–51 percent for G^C and about 7 percent for G^W. In 2020, the separation component remains large at 40.7 percent. No estimate is reported for Italy in 2018 because the corresponding LIS household file was unavailable in the estimation run. Luxembourg occupies an intermediate position: G^S remains between 28.2 and 31.1 percent, while the allocation between G^W and G^C displays somewhat greater temporal variation.

Taken together, these results indicate that the decomposition observed in 2020 is not a feature of a single cross-section. In particular, distribution-specific regional mass contributes a quantitatively meaningful share of national inequality throughout the periods.

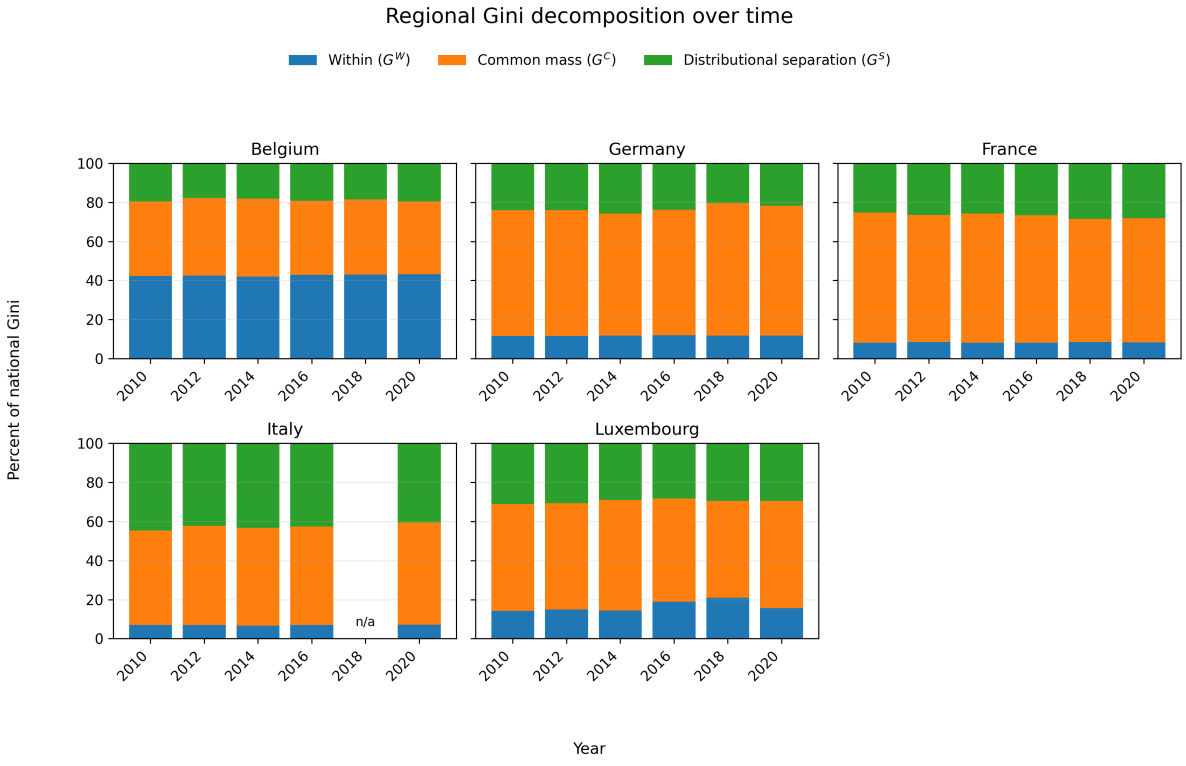


Figure 3: Regional Gini decomposition over time. Bars report the percentage contributions of the within-region component G^W, the common-mass component G^C, and the distributional-separation component G^S to the national Gini coefficient. Estimates use the baseline specification with $J = 1{,}000$ grid points and the `nrd0` bandwidth rule. Italy is unavailable in 2018 because the corresponding LIS household file could not be loaded.

4 Conclusion

This paper separates two sources of cross-group Gini dispersion. Differences in subgroup means do not fully describe differences between subgroup distributions, while the raw cross-group interaction

can remain positive even when the distributions coincide. The proposed decomposition therefore partitions that interaction into a common-mass component G^C and a distributional-separation component G^S, preserving the conventional within-group contribution so that $G = G^W + G^C + G^S$. The distinction is important for interpretation. G^C captures cross-group dispersion that can be generated entirely by distributional mass shared by the groups, whereas G^S captures dispersion involving mass that is specific to at least one group. The contribution should therefore be viewed as complementary to recent approaches that seek to define satisfactory within- and between-group components of the Gini coefficient (Attili, 2024; Heikkuri and Schief, 2026). The regional application illustrates the measurement and policy relevance of the distinction. A small conventional between-region component means that regional averages are relatively similar; it does not imply that regional distributions are approximately the same or that geography is irrelevant. Conversely, a large cross-region interaction need not imply strong regional separation when G^C is large. Mean-based decompositions should therefore not be used to dismiss the spatial dimension of inequality before the full regional distributions are examined.

References

Anderson, G., Linton, O., and Whang, Y.-J. (2012). Nonparametric estimation and inference about the overlap of two distributions. *Journal of Econometrics*, 171(1), 1–23. doi: 10.1016/j.jeconom.2012.05.001.

Akita, T. (2003). Decomposing regional income inequality in China and Indonesia using two-stage nested Theil decomposition method. *The Annals of Regional Science*, 37(1), 55–77. doi: 10.1007/s001680200107.

Attili, F. (2024). Uncovering complexities in horizontal inequality: A novel decomposition of the Gini index. *Social Indicators Research*, 173, 351–376.

Ayala, L., Martín-Román, J., and Vicente, J. (2020). The contribution of the spatial dimension to inequality: A counterfactual analysis for OECD countries. *Papers in Regional Science*, 99(3), 447–477. doi: 10.1111/pirs.12503.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica*, 47(4), 901–920. doi: 10.2307/1914138.

Breidenbach, P. and Schaffner, S. (2025). Inequality of earnings and housing costs: Developments between and within regions. *Journal for Labour Market Research*, 59(4). doi: 10.1186/s12651-024-00390-x.

Cowell, F. A. (1980). On the structure of additive inequality measures. *The Review of Economic Studies*, 47(3), 521–531. doi: 10.2307/2297303.

Dagum, C. (1997). A new approach to the decomposition of the Gini income inequality ratio. *Empirical Economics*, 22(4), 515–531.

Heikkuri, V.-M. and Schief, M. (2026). Subgroup decomposition of the Gini coefficient: A new solution to an old problem. *Econometrica*, 94(1), 169–192. doi: 10.3982/ECTA22145.

International Monetary Fund (2019). *Regional Economic Outlook: Europe*.

Lambert, P. J. and Aronson, J. R. (1993). Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, 103(420), 1221–1227. doi: 10.2307/2234247.

Novotný, J. (2007). On the measurement of regional inequality: Does spatial dimension of income inequality matter? *The Annals of Regional Science*, 41(3), 563–580. doi: 10.1007/s00168-007-0113-y.

OECD (2026). *OECD Employment Outlook 2026: Geographic Disparities in Jobs and Incomes*. Paris: OECD Publishing. doi: 10.1787/7e710f54-en.

Piacentini, M. (2014). *Measuring Income Inequality and Poverty at the Regional Level in OECD Countries*. OECD Statistics Working Papers, No. 2014/03. Paris: OECD Publishing. doi: 10.1787/5jxzf5khtg9t-en.

Pyatt, G. (1976). On the interpretation and disaggregation of Gini coefficients. *The Economic Journal*, 86(342), 243–255. doi: 10.2307/2230745.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48(3), 613–625.

Shorrocks, A. F. (1984). Inequality decomposition by population subgroups. *Econometrica*, 52(6), 1369–1385. doi: 10.2307/1913511.

Shorrocks, A. F. and Wan, G. (2005). Spatial decomposition of inequality. *Journal of Economic Geography*, 5(1), 59–81. doi: 10.1093/jnlecg/lbh054.

Silber, J. (1989). Factor components, population subgroups and the computation of the Gini index of inequality. *The Review of Economics and Statistics*, 71(1), 107–115. doi: 10.2307/1928057.

Schmid, F. and Schmidt, A. (2006). Nonparametric estimation of the coefficient of overlapping—theory and empirical application. *Computational Statistics & Data Analysis*, 50(6), 1583–1596. doi: 10.1016/j.csda.2005.01.014.

Weitzman, M. S. (1970). Measures of overlap of income distributions of White and Negro families in the United States. *U.S. Bureau of the Census, Technical Paper No. 22*. Washington, DC: U.S. Bureau of the Census.

Yitzhaki, S. (1994). Economic distance and overlapping of distributions. *Journal of Econometrics*, 61(1), 147–159. doi: 10.1016/0304-4076(94)90081-7.

Yitzhaki, S. and Lerman, R. I. (1991). Income stratification and income inequality. *Review of Income and Wealth*, 37(3), 313–329. doi: 10.1111/j.1475-4991.1991.tb00374.x.

A Proofs

A.1 Proof of theorem 1

From (3), the overall Gini coefficient is

$$G = G^W + \frac{1}{\mu} \sum_{k<\ell} w_k w_\ell \Delta_{k\ell}. \tag{32}$$

It remains to verify the pairwise identity $\Delta_{k\ell} = C_{k\ell} + S_{k\ell}$. If $0 < d_{k\ell} < 1$, the identity follows directly from the mixture expansion in (13). If $d_{k\ell} = 0$, the boundary definitions give

$$C_{k\ell} = \Delta_{k\ell}, \qquad S_{k\ell} = 0, \tag{33}$$

and if $d_{k\ell} = 1$ they give

$$C_{k\ell} = 0, \qquad S_{k\ell} = \Delta_{k\ell}. \tag{34}$$

Hence (18) holds for every pair k, ℓ. Substituting into the cross-group part of the Gini coefficient gives

$$G = G^W + \frac{1}{\mu} \sum_{k<\ell} w_k w_\ell (C_{k\ell} + S_{k\ell}) \tag{35}$$

$$= G^W + G^C + G^S. \tag{36}$$

Non-negativity follows because $G_k \geq 0$, while $C_{k\ell}$ and $S_{k\ell}$ are either non-negative absolute-distance integrals or non-negative weighted sums of such integrals. □

A.2 Relation between G^S and G^B

For $0 < d_{k\ell} < 1$, the mixture representation implies

$$\mu_k = (1 - d_{k\ell})\mu_H + d_{k\ell}\mu_P, \qquad \mu_\ell = (1 - d_{k\ell})\mu_H + d_{k\ell}\mu_Q,$$

and therefore

$$|\mu_k - \mu_\ell| = d_{k\ell}|\mu_P - \mu_Q|.$$

By Jensen's inequality,

$$A_{k\ell}^{HQ} \geq |\mu_H - \mu_Q|, \qquad A_{k\ell}^{PH} \geq |\mu_P - \mu_H|, \qquad A_{k\ell}^{PQ} \geq |\mu_P - \mu_Q|.$$

Hence, using the triangle inequality,

$$\begin{aligned} S_{k\ell} &= d_{k\ell}(1 - d_{k\ell}) \left(A_{k\ell}^{HQ} + A_{k\ell}^{PH} \right) + d_{k\ell}^2 A_{k\ell}^{PQ} \\ &\geq d_{k\ell}(1 - d_{k\ell}) \left(|\mu_H - \mu_Q| + |\mu_P - \mu_H| \right) + d_{k\ell}^2 |\mu_P - \mu_Q| \\ &\geq d_{k\ell}(1 - d_{k\ell})|\mu_P - \mu_Q| + d_{k\ell}^2 |\mu_P - \mu_Q| \\ &= d_{k\ell}|\mu_P - \mu_Q| \\ &= |\mu_k - \mu_\ell|. \end{aligned}$$

If $d_{k\ell} = 0$, then $F_k = F_\ell$, so $|\mu_k - \mu_\ell| = 0 = S_{k\ell}$. If $d_{k\ell} = 1$, then $S_{k\ell} = \Delta_{k\ell}$, and Jensen's inequality gives

$$\Delta_{k\ell} = \mathbb{E}|Y_k - Y_\ell| \geq |\mu_k - \mu_\ell|.$$

Thus,

$$S_{k\ell} \geq |\mu_k - \mu_\ell|$$

for every $d_{k\ell} \in [0, 1]$. Aggregating across pairs yields

$$G^S \geq G^B.$$

Finally, since

$$G = G^W + G^C + G^S = G^W + G^B + G^R,$$

it follows that

$$G^R = G^C + (G^S - G^B).$$

B Spatial inequality decomposition in the literature

Table 2: Selected evidence on within- and between-region income inequality

Study	**Geography**	**Method**	**Within / Between inequality**
Shorrocks and Wan (2005)	Evidence compiled from a large number of national and regional studies	Generalized entropy / spatial subgroup decompositions	Typically about **88% within / 12% between** on average in the reviewed non-exceptional cases.
Novotný (2007)	Review of 19 countries; additional comparison covering 46 countries	Theil / generalized entropy decomposition	For the comparable national studies, approximately **90% within / 10% between** on average. Country estimates vary considerably, from about 2–6% between in several high-income countries.
Akita (2003)	China and Indonesia; nested regional hierarchies	Nested Theil decomposition	Within-area inequality generally accounts for the larger part of total inequality, although the exact share depends on the spatial level used in the nested decomposition.
Piacentini (2014)	European and OECD regions	Theil-type decomposition	Approximately **94–99% within / 1–6% between** in the reported European-country examples.
International Monetary Fund (2019)	Advanced economies	Generalized entropy / Theil-L	Approximately **85–99+% within / <1–15% between**, with the between share ranging from below 1% in Austria to about 15% in Italy.
Ayala et al. (2020)	OECD countries, LIS household microdata	Mean log deviation	Typically approximately **89–99% within / 1–11% between** in the most recent observations.
Breidenbach and Schaffner (2025)	Germany; counties and 1 km^2 neighborhoods	Theil-T decomposition	Within-county inequality is the dominant component.
OECD (2026)	20 OECD countries, TL3 regions	Theil decomposition of equivalized household disposable income	Approximately **95% within / 5% between** on average.

Notes: The reported shares are not strictly comparable across studies because the underlying income concepts, inequality indices, dates, geographic units, and levels of spatial aggregation differ. "Between" generally refers to inequality associated with differences in mean outcomes across spatial groups, while "within" captures the remaining inequality within those groups. Where studies report multiple countries or years, the table reports either the authors' summary or representative ranges.

C Country Summary Statistics

Table 3: Summary statistics of national income distributions, all available years

Year	Country	Obs.	Regions	Mean	Median	SD	P10	P25	P75	P90	P90/P10	Gini
2010	Belgium	5,897	3	23,265	20,637	23,782	11,031	14,283	28,854	37,170	3.37	0.282
2010	Germany	16,672	16	21,898	19,139	14,731	9,541	13,338	26,669	36,490	3.82	0.299
2010	France	54,477	22	24,133	20,520	19,301	10,580	14,687	28,560	39,640	3.75	0.315
2010	Italy	7,903	20	17,254	15,000	11,535	7,247	10,350	21,234	28,858	3.98	0.314
2010	Luxembourg	5,440	13	41,349	36,524	23,228	19,978	26,352	50,942	67,774	3.39	0.273
2012	Belgium	6,145	3	24,805	22,265	15,385	11,754	15,265	31,171	39,636	3.37	0.278
2012	Germany	17,964	16	22,879	19,816	16,478	9,755	14,068	27,814	37,711	3.87	0.300
2012	France	56,047	22	24,886	21,029	20,871	10,750	15,175	29,200	40,710	3.79	0.320
2012	Italy	8,120	20	16,337	14,145	11,194	6,633	9,530	20,218	27,635	4.17	0.322
2012	Luxembourg	3,748	13	44,314	38,202	37,244	20,145	27,555	52,702	71,649	3.56	0.297
2014	Belgium	5,991	3	25,143	22,342	13,851	12,134	15,650	31,457	40,671	3.35	0.273
2014	Germany	15,872	16	23,572	20,489	19,181	9,933	14,441	28,722	39,178	3.94	0.302
2014	France	50,010	22	25,046	21,520	23,059	11,170	15,600	29,614	40,650	3.64	0.305
2014	Italy	8,026	20	16,846	14,714	11,641	6,986	10,000	20,855	28,196	4.04	0.314
2014	Luxembourg	3,445	13	44,034	38,742	26,774	21,095	28,225	53,702	70,909	3.36	0.279
2016	Belgium	6,033	3	25,876	23,100	15,321	12,565	16,122	32,356	41,168	3.28	0.272
2016	Germany	17,423	16	25,073	21,809	20,019	10,507	15,298	30,510	41,279	3.93	0.306
2016	France	52,093	22	25,275	21,956	18,310	11,222	15,867	29,930	40,950	3.65	0.303
2016	Italy	7,284	20	17,347	15,072	11,537	7,150	10,400	21,373	29,320	4.10	0.313
2016	Luxembourg	3,960	12	47,470	41,894	31,242	20,894	29,972	58,010	77,044	3.69	0.293
2018	Belgium	6,774	3	28,172	25,300	19,827	13,856	17,464	34,425	44,770	3.23	0.274
2018	Germany	18,619	16	27,052	23,504	24,112	11,171	16,162	32,430	43,654	3.91	0.312
2018	France	50,704	22	26,244	22,450	32,493	11,540	16,100	30,680	42,298	3.67	0.314
2018	Luxembourg	3,824	12	49,491	42,675	38,863	20,680	29,610	59,349	82,329	3.98	0.321
2020	Belgium	7,523	3	29,656	26,798	18,399	15,071	19,157	36,442	46,688	3.10	0.262
2020	Germany	12,513	16	29,132	25,419	27,858	11,616	17,450	34,955	47,782	4.11	0.320
2020	France	39,562	22	27,260	23,640	23,267	12,460	17,326	32,210	43,664	3.50	0.297
2020	Italy	6,208	20	18,108	15,605	12,611	7,026	10,278	22,901	31,319	4.46	0.328
2020	Luxembourg	4,072	13	54,292	47,299	31,079	24,892	33,984	66,565	90,703	3.64	0.281

Notes: Income is annual equivalised disposable household income. All income statistics are calculated using population weights. P10, P25, P75, and P90 denote the corresponding weighted percentiles. P90/P10 is the ratio of the 90th to the 10th percentile. Italy is unavailable in 2018 because the corresponding LIS household file could not be loaded.

D Empirical estimation procedure

The theoretical decomposition requires the group-specific income densities f_k. In the empirical application, these densities are estimated nonparametrically using the LIS population weights. Let $y_{ik} \geq 0$ denote equivalised disposable income for observation i in region k, and let a_{ik} denote its population weight. The normalized within-region weight is

$$\pi_{ik} = \frac{a_{ik}}{\sum_{j \in k} a_{jk}}.$$

Only observations with positive equivalised disposable income and positive population weights are retained in the estimation. This restriction is required because density estimation is performed on log income.

Because the income distribution is right-skewed and has positive support, the regional densities are estimated on the logarithmic income scale. Let

$$z_{ik} = \log y_{ik}.$$

For each region k, the density of log income is estimated using the weighted kernel density estimator

$$\widehat{g}_k(z) = \sum_{i \in k} \pi_{ik} \frac{1}{h_k} K\left(\frac{z - z_{ik}}{h_k}\right),$$

where K is the Gaussian kernel. The bandwidth h_k is selected separately for each region using the `nrd0` rule implemented in R. The density is then transformed back to the income scale using

$$\widehat{f}_k(y) = \frac{\widehat{g}_k(\log y)}{y}.$$

The resulting density is numerically renormalized so that it integrates to one over the evaluation grid.

Within each country, all regional densities are evaluated on the same grid of $J = 1{,}000$ points. The grid is equally spaced on the logarithmic income scale. Let $q_{0.001}$ and $q_{0.999}$ denote the 0.1st and 99.9th percentiles of the country income sample. The lower and upper endpoints are defined as

$$L = \max\left\{\min_i y_i, \frac{q_{0.001}}{2}\right\}, \qquad U = 2q_{0.999}.$$

The evaluation points are therefore

$$y_j = \exp\left[\log L + \frac{j-1}{J-1}(\log U - \log L)\right], \qquad j = 1, \ldots, J.$$

Numerical integrations over this grid are evaluated using trapezoidal quadrature.

For each pair of regions k and ℓ, the estimated common subdensity is constructed pointwise as

$$\widehat{m}_{k\ell}(y) = \min\left\{\widehat{f}_k(y), \widehat{f}_\ell(y)\right\}.$$

The estimated common-mass contribution to the pairwise cross-region interaction is then

$$\widehat{C}_{k\ell} = \int_0^\infty \int_0^\infty |x - y| \widehat{m}_{k\ell}(x) \widehat{m}_{k\ell}(y)\, dx\, dy,$$

where the integral is evaluated numerically on the common grid.

The total pairwise cross-region interaction is not estimated from the kernel densities. Instead, it is calculated directly from the weighted empirical distributions:

$$\widehat{\Delta}_{k\ell} = \sum_{i \in k} \sum_{j \in \ell} \pi_{ik} \pi_{j\ell} |y_{ik} - y_{j\ell}|.$$

The distributional-separation component is then obtained as the difference between the total cross-region interaction and the common-mass component,

$$\widehat{S}_{k\ell} = \widehat{\Delta}_{k\ell} - \widehat{C}_{k\ell}.$$

Thus, kernel density estimation is used to determine how the observed pairwise cross-region interaction is divided between common and distribution-specific mass; it does not determine the total cross-region interaction itself.

E Robustness checks

This section reports two robustness exercises and presents each result in both relative and absolute terms. Percentage-point changes and percentage shares show the economic magnitude relative to the national Gini, while the accompanying Gini-unit rows show the corresponding change or component in the original index scale.

Grid resolution. The baseline estimates use $J = 1{,}000$ evaluation points, as in Appendix D. The robustness exercise re-estimates the decomposition using $J = 500$ and $J = 2{,}000$, while holding fixed the `nrd0` bandwidth rule, the support of the common grid, the income definition, the population weights, and the regional grouping. Importantly, $\widehat{\Delta}_{k\ell}$ continues to be calculated directly from the weighted empirical distributions in all three specifications. Hence the total cross-region interaction is fixed; changing J affects only the numerical evaluation of $\widehat{C}_{k\ell}$ and therefore reallocates a very small amount between G^C and G^S.

Table 4 reports the change relative to the $J = 1{,}000$ baseline for every available country-year. For each country-year, the first row gives the change in percentage points of the national Gini, and the second row gives the same change in Gini-index units. The deviations are extremely small. With $J = 500$, the largest absolute component change is 0.047 percentage points, observed for Italy in 2016; with $J = 2{,}000$, the largest absolute change falls to 0.012 percentage points, again for Italy in 2016. In Gini units these differences are approximately 0.00015 and 0.00004, respectively. The opposite signs of the changes in G^C and G^S reflect the fact that their sum is fixed by the direct empirical estimate of the total cross-region interaction.

Table 4: Grid-resolution robustness: changes relative to the $J = 1{,}000$ baseline

			$J = 500$		$J = 2{,}000$	
Year	Country	Measure	ΔG^C	ΔG^S	ΔG^C	ΔG^S
2010	Belgium	Change (pp)	+0.008	-0.008	-0.002	+0.002
		Change (Gini)	+0.000023	-0.000023	-0.000006	+0.000006
2010	Germany	Change (pp)	+0.016	-0.016	-0.004	+0.004
		Change (Gini)	+0.000048	-0.000048	-0.000012	+0.000012
2010	France	Change (pp)	+0.022	-0.022	-0.006	+0.006
		Change (Gini)	+0.000069	-0.000069	-0.000019	+0.000019
2010	Italy	Change (pp)	+0.041	-0.041	-0.010	+0.010
		Change (Gini)	+0.000129	-0.000129	-0.000031	+0.000031
2010	Luxembourg	Change (pp)	+0.013	-0.013	-0.003	+0.003
		Change (Gini)	+0.000035	-0.000035	-0.000008	+0.000008
2012	Belgium	Change (pp)	+0.007	-0.007	-0.002	+0.002
		Change (Gini)	+0.000019	-0.000019	-0.000006	+0.000006
2012	Germany	Change (pp)	+0.017	-0.017	-0.004	+0.004
		Change (Gini)	+0.000051	-0.000051	-0.000012	+0.000012
2012	France	Change (pp)	+0.022	-0.022	-0.006	+0.006
		Change (Gini)	+0.000070	-0.000070	-0.000019	+0.000019
2012	Italy	Change (pp)	+0.026	-0.026	-0.006	+0.006
		Change (Gini)	+0.000084	-0.000084	-0.000019	+0.000019
2012	Luxembourg	Change (pp)	+0.017	-0.017	-0.004	+0.004
		Change (Gini)	+0.000050	-0.000050	-0.000012	+0.000012
2014	Belgium	Change (pp)	+0.008	-0.008	-0.002	+0.002
		Change (Gini)	+0.000022	-0.000022	-0.000005	+0.000005
2014	Germany	Change (pp)	+0.015	-0.015	-0.004	+0.004
		Change (Gini)	+0.000045	-0.000045	-0.000012	+0.000012
2014	France	Change (pp)	+0.021	-0.021	-0.005	+0.005
		Change (Gini)	+0.000064	-0.000064	-0.000015	+0.000015
2014	Italy	Change (pp)	+0.039	-0.039	-0.010	+0.010
		Change (Gini)	+0.000122	-0.000122	-0.000031	+0.000031
2014	Luxembourg	Change (pp)	+0.025	-0.025	-0.007	+0.007
		Change (Gini)	+0.000070	-0.000070	-0.000020	+0.000020
2016	Belgium	Change (pp)	+0.008	-0.008	-0.002	+0.002
		Change (Gini)	+0.000022	-0.000022	-0.000005	+0.000005
2016	Germany	Change (pp)	+0.016	-0.016	-0.004	+0.004
		Change (Gini)	+0.000049	-0.000049	-0.000012	+0.000012
2016	France	Change (pp)	+0.025	-0.025	-0.006	+0.006
		Change (Gini)	+0.000076	-0.000076	-0.000018	+0.000018
2016	Italy	Change (pp)	+0.047	-0.047	-0.012	+0.012
		Change (Gini)	+0.000147	-0.000147	-0.000038	+0.000038
2016	Luxembourg	Change (pp)	+0.021	-0.021	-0.005	+0.005
		Change (Gini)	+0.000062	-0.000062	-0.000015	+0.000015
2018	Belgium	Change (pp)	+0.008	-0.008	-0.002	+0.002
		Change (Gini)	+0.000022	-0.000022	-0.000005	+0.000005
2018	Germany	Change (pp)	+0.021	-0.021	-0.005	+0.005
		Change (Gini)	+0.000066	-0.000066	-0.000016	+0.000016
2018	France	Change (pp)	+0.024	-0.024	-0.006	+0.006
		Change (Gini)	+0.000075	-0.000075	-0.000019	+0.000019
2018	Luxembourg	Change (pp)	+0.028	-0.028	-0.007	+0.007
		Change (Gini)	+0.000090	-0.000090	-0.000022	+0.000022
2020	Belgium	Change (pp)	+0.009	-0.009	-0.002	+0.002
		Change (Gini)	+0.000024	-0.000024	-0.000005	+0.000005
2020	Germany	Change (pp)	+0.032	-0.032	-0.008	+0.008
		Change (Gini)	+0.000102	-0.000102	-0.000026	+0.000026
2020	France	Change (pp)	+0.024	-0.024	-0.006	+0.006
		Change (Gini)	+0.000071	-0.000071	-0.000018	+0.000018
2020	Italy	Change (pp)	+0.042	-0.042	-0.011	+0.011
		Change (Gini)	+0.000138	-0.000138	-0.000036	+0.000036
2020	Luxembourg	Change (pp)	+0.014	-0.014	-0.004	+0.004
		Change (Gini)	+0.000039	-0.000039	-0.000011	+0.000011

Notes: "Change (pp)" reports percentage-point changes relative to the $J = 1{,}000$ baseline. "Change (Gini)" reports the same change in Gini-index units. All other settings are held fixed. Because the total cross-region interaction $\widehat{\Delta}_{k\ell}$ is estimated directly from the weighted empirical distributions in every grid specification, changes in G^C are offset by changes in G^S up to numerical precision. Italy is unavailable in 2018.

Direct empirical versus KDE-based cross-region interaction. The second robustness exercise changes the estimator of the total pairwise cross-region interaction. The baseline follows Appendix D and calculates

$$\widehat{\Delta}_{k\ell} = \sum_{i \in k} \sum_{j \in \ell} \pi_{ik} \pi_{j\ell} |y_{ik} - y_{j\ell}|$$

directly from the weighted empirical income distributions. The alternative computes the same absolute-distance interaction from the estimated regional densities,

$$\widehat{\Delta}_{k\ell}^{KDE} = \int_0^\infty \int_0^\infty |x - y| \widehat{f}_k(x) \widehat{f}_\ell(y) \, dx \, dy.$$

The comparison is performed at $J = 1{,}000$, using the same `nrd0` bandwidth and support rule as the baseline. The common-mass estimate G^C is therefore unchanged; the alternative estimator changes the estimated total cross-region interaction and, consequently, G^S.

Table 5 reports the full comparison for every available country-year. Each country-year is shown on two rows: the first gives G^S and the total cross-region interaction as percentages of the national Gini, while the second gives the same quantities in Gini-index units. The KDE-based specification produces larger deviations than the grid-resolution exercise, as expected because it changes the estimator of $\Delta_{k\ell}$ rather than only the numerical integration resolution. Nevertheless, G^S remains positive and quantitatively relevant in every available country-year. In 2020, for example, G^S changes from 19.5 to 19.6 percent in Belgium, from 21.8 to 24.4 percent in Germany, from 28.1 to 26.7 percent in France, from 40.7 to 46.3 percent in Italy, and from 29.6 to 33.0 percent in Luxembourg.

Table 5: Robustness to the estimator of the total cross-region interaction

Year	Country	Measure	G^S_{emp}	G^S_{KDE}	Difference	Cross_{emp}	Cross_{KDE}	Reconstruction error
2010	Belgium	Share (%)	19.6	19.3	-0.3	57.7	57.4	–
		Gini value	0.0552	0.0544	-0.00080	0.163	0.162	+0.00079
2010	Germany	Share (%)	24.0	25.1	+1.1	88.4	89.5	–
		Gini value	0.0717	0.0749	+0.00320	0.264	0.267	-0.00321
2010	France	Share (%)	25.3	25.5	+0.2	91.8	92.0	–
		Gini value	0.0795	0.0801	+0.00060	0.289	0.289	-0.00058
2010	Italy	Share (%)	44.7	47.6	+2.9	93.1	96.0	–
		Gini value	0.14	0.149	+0.00900	0.292	0.302	-0.00914
2010	Luxembourg	Share (%)	31.1	34.4	+3.3	85.7	89.1	–
		Gini value	0.0846	0.0938	+0.00920	0.234	0.243	-0.00914
2012	Belgium	Share (%)	17.8	18.4	+0.6	57.5	58.1	–
		Gini value	0.0496	0.0512	+0.00160	0.16	0.162	-0.00161
2012	Germany	Share (%)	24.1	24.5	+0.4	88.4	88.8	–
		Gini value	0.0723	0.0735	+0.00120	0.265	0.266	-0.00118
2012	France	Share (%)	26.4	26.6	+0.2	91.5	91.7	–
		Gini value	0.0844	0.0851	+0.00070	0.293	0.293	-0.00067
2012	Italy	Share (%)	42.3	45.3	+3.0	93.0	96.0	–
		Gini value	0.136	0.146	+0.01000	0.299	0.309	-0.00980
2012	Luxembourg	Share (%)	30.8	30.0	-0.8	85.1	84.4	–
		Gini value	0.0914	0.0892	-0.00220	0.253	0.251	+0.00222
2014	Belgium	Share (%)	18.2	19.3	+1.1	58.0	59.1	–
		Gini value	0.0496	0.0527	+0.00310	0.158	0.161	-0.00311
2014	Germany	Share (%)	25.7	25.5	-0.2	88.2	88.0	–
		Gini value	0.0776	0.077	-0.00060	0.266	0.266	+0.00053
2014	France	Share (%)	25.7	25.1	-0.6	91.9	91.3	–
		Gini value	0.0785	0.0767	-0.00180	0.28	0.279	+0.00176
2014	Italy	Share (%)	43.3	45.1	+1.8	93.4	95.3	–
		Gini value	0.136	0.142	+0.00600	0.293	0.299	-0.00587
2014	Luxembourg	Share (%)	29.0	32.8	+3.8	85.6	89.4	–
		Gini value	0.0811	0.0916	+0.01050	0.239	0.25	-0.01050
2016	Belgium	Share (%)	19.2	19.8	+0.6	57.1	57.8	–
		Gini value	0.0521	0.0539	+0.00180	0.155	0.157	-0.00182
2016	Germany	Share (%)	23.9	24.5	+0.6	88.1	88.6	–
		Gini value	0.0731	0.0748	+0.00170	0.269	0.271	-0.00170
2016	France	Share (%)	26.6	27.1	+0.5	91.8	92.3	–
		Gini value	0.0805	0.082	+0.00150	0.278	0.28	-0.00148
2016	Italy	Share (%)	42.7	45.6	+2.9	93.0	96.0	–
		Gini value	0.133	0.143	+0.01000	0.291	0.3	-0.00929
2016	Luxembourg	Share (%)	28.2	30.6	+2.4	81.1	83.4	–
		Gini value	0.0828	0.0896	+0.00680	0.238	0.244	-0.00678
2018	Belgium	Share (%)	18.6	18.4	-0.2	57.0	56.8	–
		Gini value	0.0509	0.0504	-0.00050	0.156	0.156	+0.00050
2018	Germany	Share (%)	20.4	23.6	+3.2	88.2	91.4	–
		Gini value	0.0638	0.0737	+0.00990	0.276	0.286	-0.00996
2018	France	Share (%)	28.7	26.6	-2.1	91.5	89.4	–
		Gini value	0.0899	0.0833	-0.00660	0.287	0.28	+0.00655
2018	Luxembourg	Share (%)	29.5	32.1	+2.6	79.1	81.6	–
		Gini value	0.0947	0.103	+0.00830	0.254	0.262	-0.00808
2020	Belgium	Share (%)	19.5	19.6	+0.1	56.8	56.9	–
		Gini value	0.051	0.0514	+0.00040	0.149	0.149	-0.00039
2020	Germany	Share (%)	21.8	24.4	+2.6	88.3	91.0	–
		Gini value	0.0697	0.0781	+0.00840	0.282	0.291	-0.00840
2020	France	Share (%)	28.1	26.7	-1.4	91.6	90.2	–
		Gini value	0.0834	0.0792	-0.00420	0.272	0.268	+0.00425
2020	Italy	Share (%)	40.7	46.3	+5.6	92.9	98.5	–
		Gini value	0.133	0.152	+0.01900	0.305	0.323	-0.01840
2020	Luxembourg	Share (%)	29.6	33.0	+3.4	84.4	87.8	–
		Gini value	0.0831	0.0927	+0.00960	0.237	0.247	-0.00962

Notes: "Share (%)" reports percentages of the national Gini; its Difference column is in percentage points. "Gini value" reports the corresponding quantities in Gini-index units; its Difference column is $G^S_{KDE} - G^S_{emp}$. G^S_{emp} uses the direct weighted empirical estimator of $\Delta_{k\ell}$ and corresponds to the paper baseline. G^S_{KDE} uses the KDE-based approximation. "Cross" is the total cross-region interaction. The reconstruction error is $G - (G^W + G^C + G^S_{KDE})$ and is therefore reported only in Gini units. Italy is unavailable in 2018.

Across the available years, the largest absolute change in G^S is 1.1 percentage points for Belgium, 3.2 for Germany, 2.1 for France, 5.6 for Italy, and 3.8 for Luxembourg. Table 6 reports these maxima

in both percentage points and Gini units.

Table 6: Largest change in G^S under the KDE-based interaction estimator

Country	Change in G^S (pp)	Change in G^S (Gini)	Year
Belgium	+1.1	+0.00310	2014
Germany	+3.2	+0.00990	2018
France	-2.1	-0.00660	2018
Italy	+5.6	+0.01900	2020
Luxembourg	+3.8	+0.01050	2014

Notes: For each country, the table reports the country-year with the largest absolute difference between the KDE-based and direct empirical estimates of G^S. The signs indicate the direction of the KDE-based change.

The direct empirical estimator remains the preferred specification. With the direct estimator, $G^W + G^C + G^S$ reproduces the observed national Gini up to numerical precision because the empirical $\Delta_{k\ell}$ measures the total observed pairwise cross-region interaction directly. When $\Delta_{k\ell}$ is instead approximated from smoothed densities, exact reconstruction is no longer guaranteed. The largest discrepancy in the reported results occurs for Italy in 2020, where the KDE-based specification raises G^S from 40.7 to 46.3 percent (from approximately 0.133 to 0.152 Gini points), and the reconstructed Gini differs from the observed Gini by approximately 0.018 Gini points. The robustness exercises therefore support two conclusions: the baseline decomposition is numerically stable with respect to the integration grid, and the substantive importance of distributional separation remains under a materially different, fully density-based estimator of the cross-region interaction.

Different bandwidth As a sensitivity check, the regional kernel densities are re-estimated using bandwidths equal to $0.8h_{\text{nrd0}}$, h_{nrd0}, and $1.2h_{\text{nrd0}}$, while all other estimation choices are held fixed. The national Gini G and the within-region contribution G_W are unchanged across bandwidth specifications. Bandwidth variation therefore affects only the allocation of the cross-regional component between the common component G_C and the separation component G_S. A narrower bandwidth lowers G_C and raises G_S, whereas a wider bandwidth produces the opposite pattern. The changes are small for Belgium and France and somewhat larger for Germany, Italy, and Luxembourg. Importantly, the total cross-regional contribution $G_C + G_S$ remains unchanged.

Table 7: Bandwidth robustness of the Gini decomposition

Country	Bandwidth	G	G_W	G_C	G_S
Belgium	$0.8h_{\text{nrd0}}$	0.274	0.117 (42.6%)	0.104 (38.0%)	0.053 (19.4%)
	h_{nrd0}	0.274	0.117 (42.6%)	0.105 (38.5%)	0.051 (18.8%)
	$1.2h_{\text{nrd0}}$	0.274	0.117 (42.6%)	0.107 (39.1%)	0.050 (18.2%)
France	$0.8h_{\text{nrd0}}$	0.309	0.026 (8.3%)	0.197 (63.7%)	0.086 (28.0%)
	h_{nrd0}	0.309	0.026 (8.3%)	0.200 (64.9%)	0.083 (26.8%)
	$1.2h_{\text{nrd0}}$	0.309	0.026 (8.3%)	0.204 (66.0%)	0.079 (25.7%)
Germany	$0.8h_{\text{nrd0}}$	0.306	0.036 (11.7%)	0.193 (63.0%)	0.077 (25.2%)
	h_{nrd0}	0.306	0.036 (11.7%)	0.199 (65.0%)	0.071 (23.3%)
	$1.2h_{\text{nrd0}}$	0.306	0.036 (11.7%)	0.205 (66.8%)	0.066 (21.4%)
Italy	$0.8h_{\text{nrd0}}$	0.318	0.022 (6.9%)	0.153 (48.2%)	0.143 (44.9%)
	h_{nrd0}	0.318	0.022 (6.9%)	0.160 (50.4%)	0.136 (42.7%)
	$1.2h_{\text{nrd0}}$	0.318	0.022 (6.9%)	0.167 (52.6%)	0.129 (40.5%)
Luxembourg	$0.8h_{\text{nrd0}}$	0.291	0.048 (16.5%)	0.150 (51.6%)	0.093 (31.9%)
	h_{nrd0}	0.291	0.048 (16.5%)	0.156 (53.8%)	0.086 (29.7%)
	$1.2h_{\text{nrd0}}$	0.291	0.048 (16.5%)	0.162 (55.9%)	0.080 (27.6%)

Notes: Entries for G_W, G_C, and G_S report the absolute Gini contribution, with its share of the national Gini in parentheses. Values are averages across the available country-year estimates. The bandwidth robustness varies the region-specific `nrd0` bandwidth by $\pm 20\%$, using $0.8h_{\text{nrd0}}$, h_{nrd0}, and $1.2h_{\text{nrd0}}$, while all other estimation choices are held fixed. Italy is averaged over the five years available in the robustness output; the other countries use six years.