BANCA D'ITALIA

EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

The geography of Italian income inequality: recent trends and the role of employment

by Emanuele Ciani and Roberto Torrini

April 2019

Number 492

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

The geography of Italian income inequality: recent trends and the role of employment

by Emanuele Ciani and Roberto Torrini

Number 492 – April 2019

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

THE GEOGRAPHY OF ITALIAN INCOME INEQUALITY: RECENT TRENDS AND THE ROLE OF EMPLOYMENT

by Emanuele Ciani* and Roberto Torrini*

Abstract

We reassess the role of regional imbalances in explaining the high household income inequality in Italy. In the first part of the work we use the Survey of Household Income and Wealth (SHIW) to describe the trends in income inequality between and within areas since the early 2000s. We illustrate that the between-area inequality has been relatively stable, while the within-area component increased significantly after the recession and during the recovery. In 2016, the large geographical divide and the higher inequality within the South contributed to almost one fifth of national inequality. In the second part we show that the distribution of employment is key in explaining the regional differences in both average income and its dispersion. By means of simulations based on matching and reweighting, we estimate that national inequality would be reduced by 15 per cent if the distribution of work hours across southern households was similar to the one in the more developed Centre-North. Regional employment differentials are so important in determining overall inequality that income dispersion would decline substantially even if this increase in employment was associated with a drop in southern regions' average wages.

JEL Classification: D63, R10, J21.
Keywords: inequality, North-South divide, work intensity.

Contents

* Bank of Italy, Directorate General for Economics, Statistics and Research.

1. Introduction[1]

Household income inequality in Italy is comparatively high. This is due to both high market income inequality and a poorly designed tax and benefit system. In particular, social expenditure is quite generous to retired workers, but does little to support young households whose members are temporarily or permanently out of employment (panel (a) of Figure 1). These limitations of the Italian welfare system are well-known and extensively discussed in the literature (starting with Gorrieri, 1972; for a more recent analysis see Baldini et al., 2002 and Baldini and Toso, 2009). Recommendations for improving its effectiveness have become a common feature of the international institutions' reports on the Italian economy (e.g. European Commission, IMF and OECD). Less effort has been devoted to understanding the sources of market income inequality and the role played by the pronounced regional disparities.

The North-South divide is, however, key to understanding equivalent income inequality in Italy and to designing policies aimed at achieving a more equitable income distribution.[2] As shown by Brandolini and Torrini (2010), regional disparities explain a large share of income dispersion. Their impact is larger than in Germany (see panel (b) of Figure 1) and Spain, where geographical heterogeneity is also relevant. The contribution of regional imbalances to overall household income inequality is twofold: southern regions are characterized not only by a lower household income, but also by a much higher within-area inequality. The Gini index for the South was almost four percentage points higher than in the Centre-North in 2016. In the first part of the work, using data from the Survey of Household Income and Wealth (SHIW), we reassess the importance of regional disparities, analysing the contribution of the between- and within-area components of household income inequality. Looking at the mean log deviation, an inequality index that can be decomposed in the two components, we show that the difference in average income between the South and the Centre-North accounted for 12 per cent of overall income dispersion in 2016. The within-area inequality in the South (equal to 0.216) was significantly higher than in the Centre-North (0.171). Reducing the southern regions' inequality to the levels observed in the Centre-North would shrink national income inequality from 0.212 to 0.197 (7 per cent). If we also eliminated the gap in average income inequality between the two areas, Italian income inequality would be further reduced to 0.171.

Looking at the dynamics of the different components of income inequality we also show that, while the between-area contribution was relatively stable between 2000 and 2016, the within-area inequality increased significantly after the recession and during the ensuing recovery. The economic recession hit the bottom of the income distribution more severely: low income households

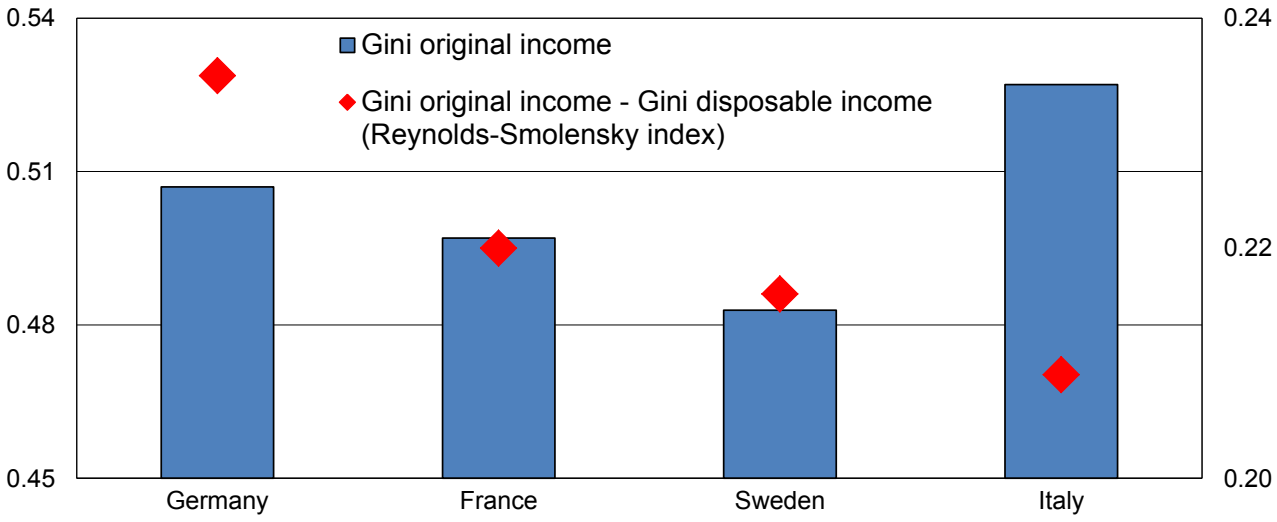
[1] The views expressed in this paper are those of the authors and do not necessarily correspond to those of the Institution to which they are affiliated.

[2] Equivalent income is the ratio between total household income and the number of equivalent adults. The latter is calculated using the OECD-modified equivalence scale which assigns a value of 1 to the head of household, of 0.5 to each additional member over the age of 14 and of 0.3 to each child below that age. The unit of reference is the individual. Estimates are calculated using sample weights; values are revalued as at 2016 using the consumption deflator for resident households published by Istat in the national accounts. In the paper the terms 'income' and 'equivalized household income' will be used interchangeably, unless we refer explicitly to specific components (such as labour income). The sources of data are always the Bank of Italy's Survey on Household Income and Wealth and the Historical Archives (version 10.0, January 2018). Only for the calculation of the mean logarithmic deviation were incomes below the second percentile of the distribution moved up to that percentile (bottom-coding), separately for each year, to prevent extremely low values from skewing the results.

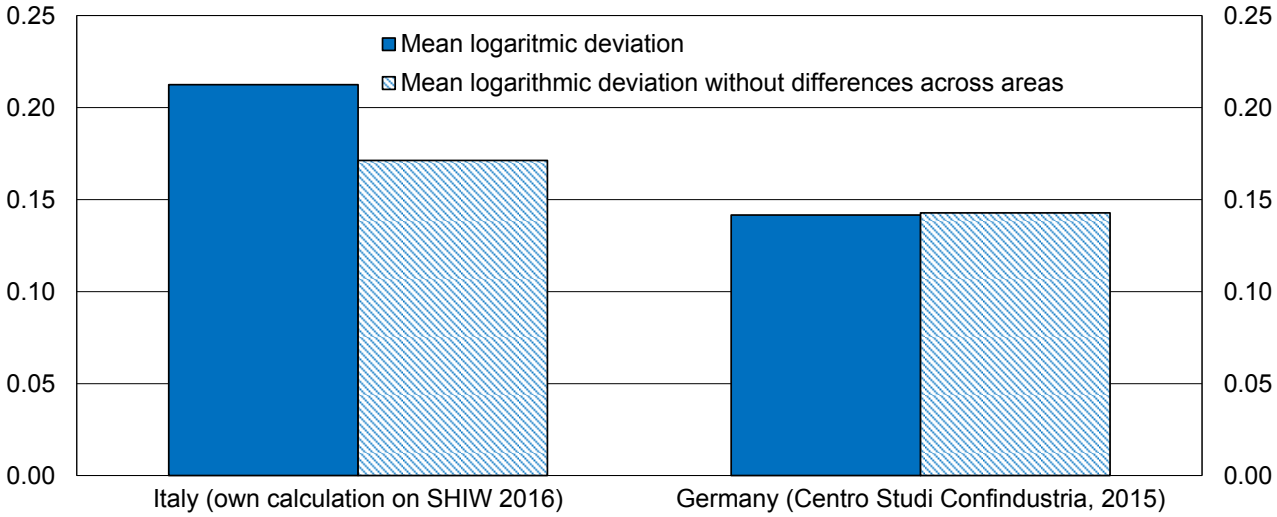
experienced a strong contraction in both labour and transfer income, especially in the South. The increase in the foreign-born population accounts for a large part of the increase in inequality in the Centre-North. Changes in this area in the last decade are negligible if we only consider households whose reference person is born in Italy, while the trends in the South remain similar.

Figure 1. Inequality in Italy and other countries

(a) Gini index of original income (before tax and transfers) vs Gini index of disposable income (after tax and transfers); Euromod, 2016.



(b) Inequality and differences across areas



Note: In panel (b), the two areas in Italy are Centre-North and South; in Germany they refer to the Länder in the former West and East of Germany. To calculate the mean logarithmic deviation without differences, the between-area component is set to zero and the inequality in the poorer area is brought to the levels of the richer one. For Germany the increase in overall inequality is due to the fact that the East of Germany has lower within inequality than the West.

Having assessed the importance of regional imbalances in explaining income inequality in Italy, the main contribution of the paper is to show that the distribution of employment opportunities is crucial in explaining the within- and between-area income distribution. For this purpose, in the second part of the work we discuss univariate decompositions of income distribution with respect to household demographic and socio economic characteristics. According to a simple counterfactual-

accounting exercise based on these decompositions, if we were able to increase employment intensity in southern households to the levels observed in the Centre-North, overall inequality would drop by 12 per cent, from 0.212 to 0.187 (assuming that the wage structure had remained the same in each area); income dispersion in the South would drop from 0.216 to 0.185.

To complement these univariate decompositions we discuss counterfactual simulations based on matching methods, which we use to assign the labour market outcomes (work intensity and/or earnings) of households in the Centre-North with similar observable characteristics to each southern household. We complement this matching-based simulation with a reweighting exercise (Fortin et al., 2011). All the results suggest that both within- and between-area would decrease following the growth in work intensity, with a drop in overall inequality by around 15 per cent.

These accounting exercises suggest that tackling the lack of work opportunities in southern regions is key to addressing the problem of inequality in Italy. This does not mean that redistributive policies are irrelevant, nor does it imply that the country can do without a well-targeted tax and benefit system. However, more generous redistributive policies should be carefully designed in order to avoid introducing disincentives to work, employment being a key component of the problem to be solved.[3]

How to increase employment levels in southern regions is nevertheless an open question. Policies that favour labour supply, such as introducing in-work benefits and reducing the tax wedge on low incomes, could be a positive step towards reducing the structurally low southern employment rate. Nevertheless, labour demand in the South is affected by the lower labour productivity that characterizes the area. Increasing productivity is therefore an obvious necessity. However, the experiments of the last sixty years, starting with the *Cassa per il Mezzogiorno* and continuing with the European cohesion policy, have shown a limited ability of place-based policies to fill the gap with respect to the rest of the country (Banca d'Italia, 2010). Furthermore, policies aimed at rising average productivity in the South might take a long time before they start to pay off. A complementary policy, which might produce the desired effects also in the short term, could be to decrease labour costs in the South to stimulate employment growth. This could be obtained by reducing the wedge between labour cost and take-home pay (for instance by subsidizing the fraction of social contributions paid by the employer[4]) or by reforming the centralized bargaining in order to allow more pronounced regional wage differentials (Bodo and Sestito, 1991; Faini, 1999; Accetturo et al., 2009; Boeri et al., 2019).[5] In the latter case, the increase in employment might come alongside a contraction in net hourly labour earnings. This contraction would at least partially offset the reduction in the between-area income gap associated with a rise in employment in Southern regions. We abstain here from considerations about which size of the cut would be necessary to increase employment in the South at the levels observed in the Centre-North: this would require the calibration of a spatial equilibrium model. We nevertheless simulate the impact on inequality of a rise in the employment intensity in the area associated with net labour earning cuts of different

[3] These policies might use either a homogeneous means-testing method across the country, or set region-specific thresholds. However, also in the case of a homogeneous mean-testing method, most of the resources are likely to flow towards the poorer southern households.

[4] However, European rules on State aid limit the possibility of implementing a generalised reduction of tax and social contributions on a permanent basis at the regional level; see European Commission (2014).

[5] There is evidence that centralized bargaining limits the variability of wages across Italian regions (Bodo and Sestito, 1991; Boeri et al., 2019; Ciani et al., 2019).

magnitude. Although this is a rough exercise, it is useful for providing further insights into the importance of work intensity in Southern regions. Only a cut in net hourly labour earnings of around 40 per cent would compensate the effect on inequality of an increase in working hours. A rise in employment intensity to the level observed in the Centre-North accompanied by a 20 per cent decline in net average labour earnings (roughly equal to the productivity gap of the area) would still imply a 10 per cent reduction in overall income inequality.

Section 2 discusses recent trends in income inequality, decomposing it in the two components of between- and within-areas. Section 3 starts by using univariate decompositions to highlight that the higher inequality within the South is mostly due to the low work intensity. It then discusses results from the different simulations. Section 4 concludes.

2. Recent trends in income inequality between and within areas

2.1 Trends in household income and its distribution

Between 2000 and 2006 the average real household income grew at a similar rate in the Centre-North and in the South (panel (a) of Figure 2). A significant contribution came from labour income, thanks to the improvement in the labour market. A smaller fraction of growth was due to the increase in pensions and other transfers (Table 1).

Table 1. Growth (per cent) in equivalent household income and contributions of its different sub-components, 2000-2016

	South				Centre-North			
		Contribution of the subcomponent:				Contribution of the subcomponent:		
	Income growth	- from work	- from transfers	- from capital	Income growth	- from work	- from transfers	- from capital
2000-06	8.2	6.3	1.5	0.4	8.1	6.3	2.0	-0.1
2006-16	-10.4	-9.0	-0.3	-1.2	-11.5	-9.1	0.9	-3.3

Note: Transfers include pensions and other transfers; capital income includes imputed rents.

Table 2. Growth (per cent) in the number of household members receiving labour income and in the average equivalent labour income per worker, 2006-2016

	Labour income per worker	Number of labour income recipients
Centre-North	-10.1	-7.4
South	-6.7	-8.6

Note: by worker here we mean a recipient of labour income. The labour income per worker is calculated considering only households with at least one labour income recipient and is calculated using equivalent income, consistently with the rest of the paper.

In the following decade, particularly during the economic recession, household income dropped, slightly more in the Centre-North because of a stronger contraction in capital income (including imputed rents).[6] The contraction in labour income was similar in the two areas. The average number

[6] This is true with respect to both the percentage variation in capital income and its contribution to overall income growth.

of employed household members decreased more in the South, but the labour income per worker (in equivalent terms) shrank more in the Centre-North (Table 2). Considering the overall period, the income gap between the South and the Centre-North has been relatively stable: it was 40 per cent in 2000, it dropped to 35 per cent in 2012-2014, but it then rose to 39 per cent in 2016.[7]

Income inequality is structurally higher within the South, according to both the Gini index and the mean logarithmic deviation (panels (a) and (b) of Figure 3). Dispersion decreased significantly in southern regions between 2000 and 2006 (by 3.6 percentage points for the Gini index and by 4.5 for the mean logarithmic deviation), but it returned to growth during the crisis and the following recovery (see Acciari and Mocetti, 2013, for an empirical analysis based on tax-income data at the provincial level).

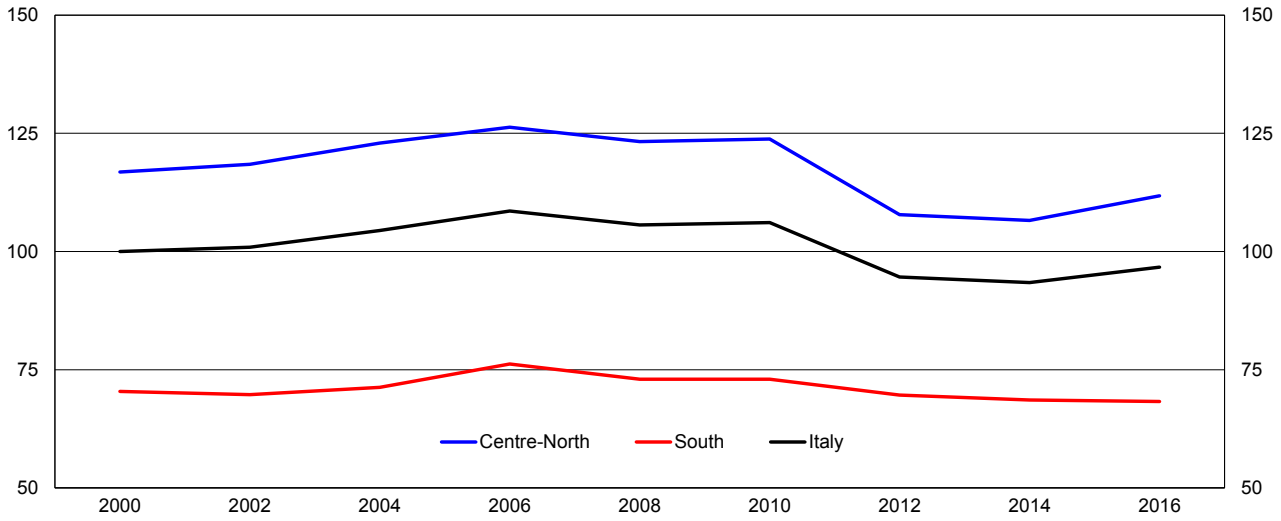
In the Centre-North, inequality remained relatively stable up to 2006 but grew in the following decade. The change in central and northern regions was partially driven by the growth in the foreign-born population, a group characterized – within the area – by lower income and higher inequality. Between 2006 and 2016 the Gini index for households with a reference person (i.e. the recipient of the highest income) born in Italy did not change in the Centre-North, while it increased in the South, similarly to the index computed for the whole population; using the mean logarithmic deviation we still observe a small increase in inequality in the Centre-North, but smaller than in trends including the foreign-born population.

By focusing on the mean logarithmic deviation, we can decompose the national inequality into a within-area component, which accounts for the dispersion of income around the average income inside each area, and a between-area one, which reflects the average income gap between the two areas (panel (c) of Figure 3). The between-area component was relatively stable during the period in question; it only shrank slightly in more recent years. Nevertheless, this component still provides a sizeable contribution, accounting for 12 per cent of overall national income inequality in 2016. The contribution of inequality within the South was decreasing before 2008 but it started growing from the recession onwards. The contribution of the Centre-North grew over time, owing not only to increasing inequality but also to the rising share of the population living in the area.

[7] If we limit the comparison to households with a native-born reference person, the income gap between the two areas increased slightly, from 40.2 per cent in 2000 to 41.0 per cent in 2006 and 41.8 per cent in 2016.

Figure 2. Equivalent household income, 2000-2016

(a) Average income (Italy 2000=100)



(b) Income gap between South and Centre-North (values and percentage points)

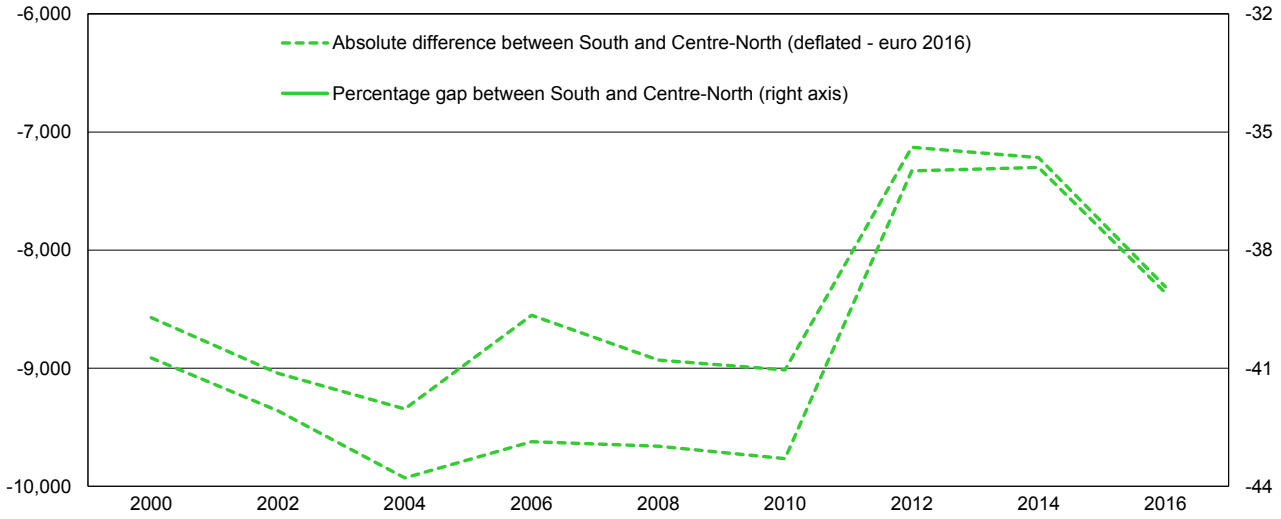
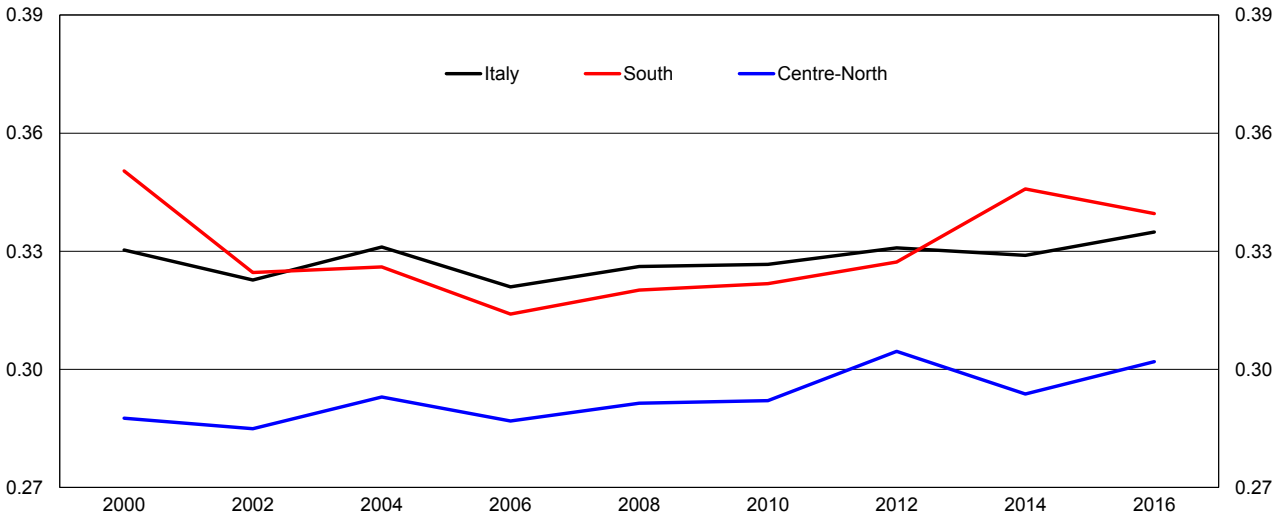
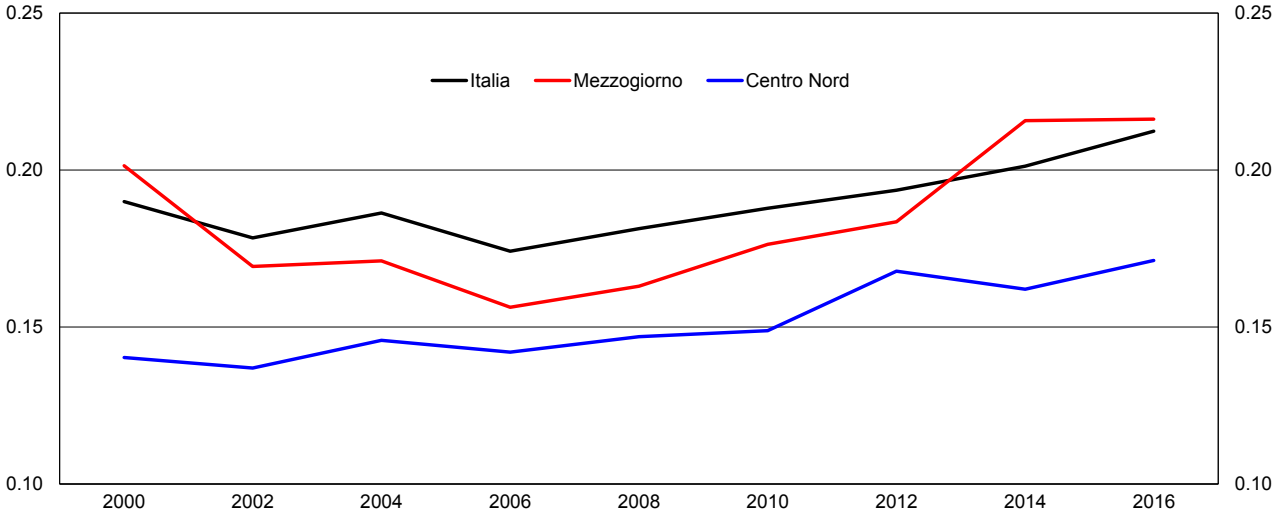


Figure 3. Inequality indices, 2000-2016

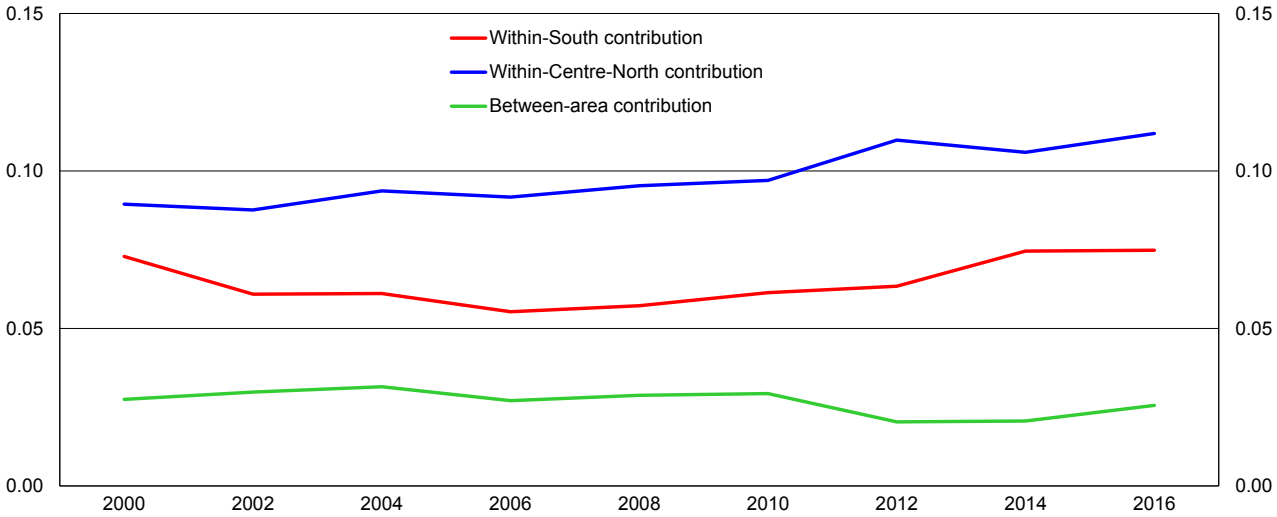
(a) Gini index (values)



(b) Mean logarithmic deviation (values)



(c) Decomposition of the mean logarithmic deviation into within- and between-area components (values)



2.2 Different trends at different percentiles?

Although the two areas displayed similar changes in average income, there are significant geographical differences at different points in the income distribution (see also Peragine, 2017).

The decrease in inequality within the South between 2000 and 2006 had been driven by the marked increase in the lowest household income (the 10th percentile of the distribution; see panel (a) of Figure 4), while higher incomes had been more stable (median and 90th percentile). However, the ensuing economic recession hit the bottom of the income distribution more severely, which experienced a strong contraction in both labour and transfer income (Table 3). The contraction between 2006 and 2016 was greater for lower income households in the Centre-North too, but the drop was smaller (-34 per cent in the average income of the first 10th of the distribution vs -56 per cent in southern regions). The geographical differences in the impact of the crisis on the bottom part of the distribution are even larger if we focus only on households whose reference person was born in Italy (-19 per cent in the Centre-North vs -55 per cent in the South).[8]

Table 3. Growth (per cent) in equivalent household income and contributions of its different sub-components at different percentiles, 2000-2016

	South				Centre-North			
		Contribution of the component:				Contribution of the component:		
	Income growth	- from work	- from transfers	- from capital	Change in income	- from work	- from transfers	- from capital
	(a) Households in the first 10th of the distribution							
2000-06	69.1	23.4	37.4	8.3	5.3	0.7	5.5	-0.9
2006-16	-56.0	-28.5	-26.0	-1.5	-33.8	-17.0	-12.9	-3.9
	(b) Households in the first fifth of the distribution							
2000-06	36.4	15.6	17.6	3.2	7.6	2.3	3.4	1.8
2006-16	-34.8	-17.6	-15.9	-1.3	-26.6	-11.5	-8.9	-6.2
	(c) Total							
2000-06	8.2	6.3	1.5	0.4	8.1	6.3	2.0	-0.1
2006-16	-10.4	-9.0	-0.3	-1.2	-11.5	-9.1	0.9	-3.3

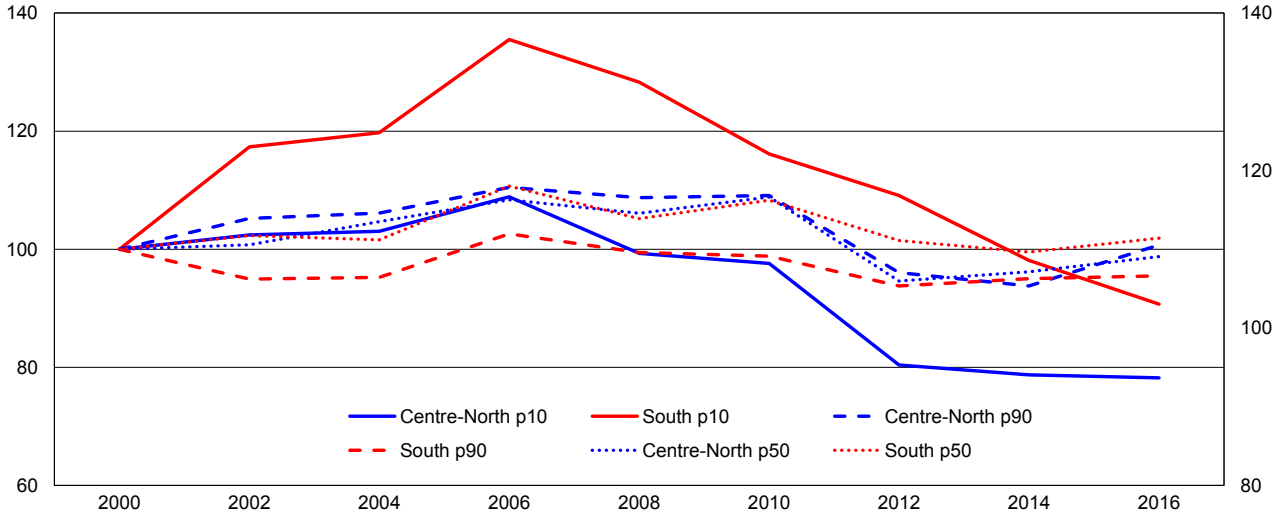
Note: Transfers include pensions and other transfers; capital income includes imputed rents. The groups (e.g. the first 10th of the distribution) are defined according to their position in the distribution within each area.

During the recession and the later recovery the geographical differences between low-income households increased. The gap between the 10th percentile in the South and the same percentile in the Centre-North, after decreasing by 10 percentage points between 2000 and 2006 to 46 per cent, grew to 49 per cent in 2016. The gap between the two areas in median income instead remained approximately the same (around 41 per cent in both 2006 and 2016), while the gap evaluated at the 90th percentile decreased slightly (from 35 per cent in 2006 to 34 per cent in 2016).

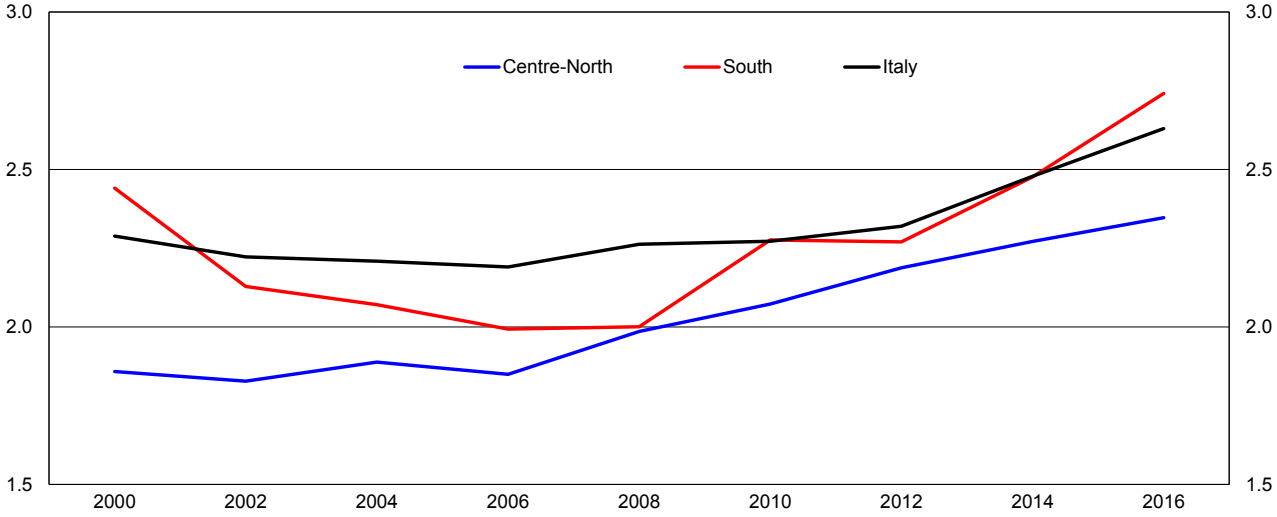
[8] In this case we focus on the first decile of the income distribution (within area) among households whose reference person was born in Italy.

Figure 4. Trend in different percentiles of the equivalent income distribution, 2000-2016

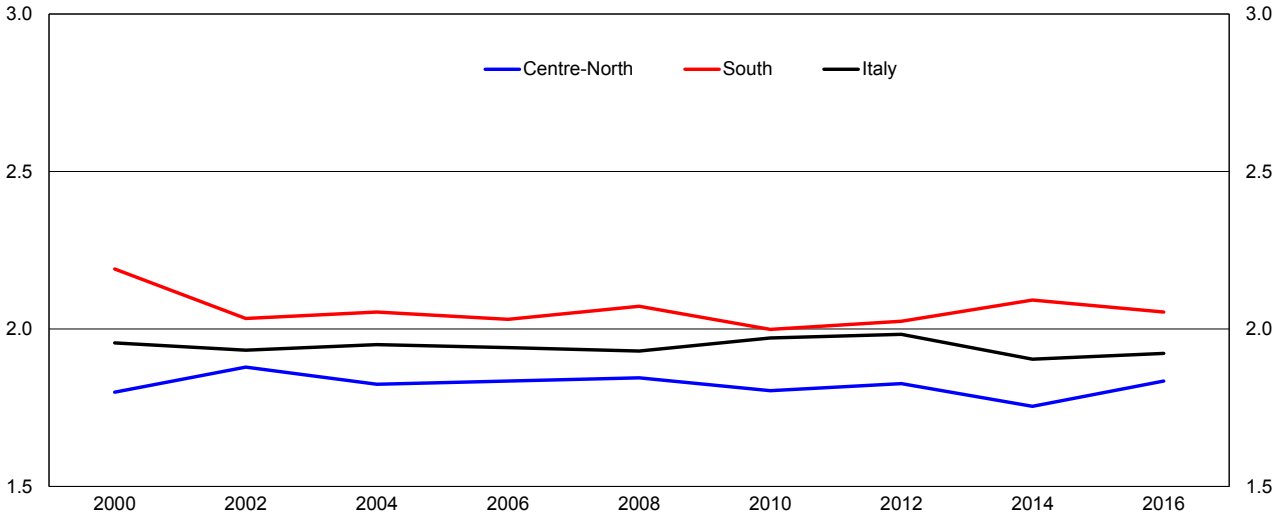
(a) Trend in the different percentiles (2000=100)



(b) Ratio between the 50th and 10th percentile (values)



(c) Ratio between the 90th and 50th percentile (values)



In 2016, 10 per cent of households at the bottom of the distribution had, on average, one third of the income of households in the corresponding group in the Centre-North, while for those in the top 10 per cent the ratio was two thirds. Considering only households with a native-born reference person, the gap between the two areas for the 10th percentile grew from 49 per cent in 2006 to 59 per cent in 2016, while in this case too, the gaps evaluated at the median and at the 90th percentiles did not change much.

Income distribution in the South is, therefore, characterized by a greater gap between low income-households and the rest of the population (panels (b) and (c) of Figure 4,), while the dispersion at the top of the distribution is more similar to the one observed in the Centre-North. The trend of inequality within the South between 2000 and 2016 seems to have been mostly influenced by the trends in the lowest percentiles.

3. The determinant of greater income inequality in the South

3.1 Southern households' characteristics

Households in the two areas are characterized by marked differences in socio-demographic characteristics and employment rates (Table 4; more details can be found in Table A1 in the Appendix), which may impact on both the level and the dispersion of income.

Table 4. Average households' socio-demographic characteristics, 2016

	Position in the income distribution					
	All		Bottom decile		Top decile	
	Centre-North	South	Centre-North	South	Centre-North	South
Reference person's age	54.5	55.2	49.9	46.0	57.8	57.1
Household size	3.0	3.3	4.0	3.7	2.7	2.8
Fraction of reference persons with a university degree (%)	16.1	9.9	5.6	0.2	55.4	41.7
Fraction living in own property (%)	71.4	67.5	27.7	35.2	93.0	88.1
Fraction of reference persons born abroad (%)	12.7	4.0	51.8	7.6	4.6	0.2
Number of income recipients	1.8	1.6	1.3	1.2	2.0	2.1
Number of labour income recipients	1.2	1.0	0.8	0.5	1.4	1.4
Fraction of adults recipients of labour income (%)	54.8	39.3	35.7	16.9	64.7	61.4
Fraction of adults recipients of labour income or work-related pension (%)	74.8	53.3	39.3	17.0	89.0	84.4
Work intensity (1)	66.4	45.0	32.2	14.6	80.0	77.6

Note: Averages are calculated by weighting households by their size. Deciles are calculated on the household income distribution within areas. (1) Only households with members aged 18-64. Work intensity is calculated as the total number of months worked by members aged 18-64 divided by the total number of months available to them (12 months times their number).

In the Centre-North the reference person is younger (54.5 years vs 55.2 in the South), more likely to hold a university degree (16.1 vs 9.9 per cent) and to have been born abroad (12.7 vs 4.0 per cent). Households are larger in southern regions (3.0 in the Centre-North vs 3.3 in the South).

There are also significant differences in the number of income recipients, in particular in the fraction of adults with labour income, which is lower in the South (39.3 vs 54.8 per cent in the Centre-North). The differences in employment-related variables are even larger if we do not only consider labour income but also the receipt of work-based pensions.[9] A similar conclusion is reached if we look at a measure of work intensity, built by looking at the fraction of months actually worked by household members aged 18-64. The differences in employment rates and work intensity are particularly strong in the decile with the lowest income, while they are small in the top one.

3.2 Which characteristics contribute more to higher inequality in the South?

To evaluate how these differences contribute to explaining the greater inequality in the South we decompose the mean logarithmic deviation into socio-demographic and economic variables. Formally, for each characteristic we define *K* approximately homogeneous groups (e.g. age classes) and we decompose the within-area index into two further components: one within groups and another between groups. Those variables that, in the South, are more important in explaining inequality should be those with a higher between-group component.

Considering the *n* individuals living in the South, let y_i be their equivalent household income and μ^S the average. The mean logarithmic deviation in the South is

$$L^S = -\frac{1}{n}\sum_{i=1}^{n}\log(\frac{y_i}{\mu^S}) \qquad (1)$$

This (within-area) index can be further decomposed into two groups, one *intra-group* and the other *inter-group*:

$$L^S = L^S_{INTRA} + L^S_{INTER} = \sum_{k=1}^{K} w_k^S L_k^S - \frac{1}{K}\sum_{k=1}^{K} w_k^S \log(\frac{\mu_k^S}{\mu^S}) \qquad (2)$$

where w_k^S is the share of the southern population belonging to the k-th group, while L_k^S is the mean logarithmic deviation in this group (in the South). Similarly we can decompose the mean logarithmic deviation into the Centre-North.

In the South, the most important inter-group components are observed for the education level (21 per cent of the overall index; Table A2 in the Appendix) and for the employment-related variables, such as the fraction of adult recipients of labour income or a work-based pension (26 per cent) and the work intensity measure combined with the presence of pension recipients (26 per cent). For these variables the difference with the inter-group component measured for the Centre and North is sizeable. Furthermore, the population in the South is concentrated into groups, such as those with low work intensity, which are associated with a larger within-group inequality.

Following Brandolini and D'Alessio (2001) and Brandolini (2009), this decomposition can be used to perform some counterfactual exercises, in which we study how inequality would change if the

[9] By work-based pensions we refer to *pensioni di anzianità* and *pensioni di vecchiaia.*

distribution of the southern population across the different groups is changed to mimic that observed in the Centre-North. If the higher inequality in the South is due to differences in a particular characteristic, then the mean logarithmic deviation should drop when we reweight the population shares to make southern households more similar to those living in the Centre-North with respect to this characteristic.

For each dimension, the share of southern residents in group *k* is assumed to be equal to the one in the Centre-North, while the income structure (both the average income and the within- inequality of each group) is left unchanged. The new index with adjusted population shares is:

$$L_A^{MZ} = \sum_{k=1}^{K} w_k^{CN} L_k^{MZ} - \frac{1}{K}\sum_{k=1}^{K} w_k^{CN} \log\left(\frac{\mu_k^{MZ}}{\mu_A^{MZ}}\right) \qquad (3)$$

where the average income in the South μ_A^{MZ} is recalculated using the new population shares. If the differences across areas in a specific characteristic contribute to explaining the higher inequality in the South, then we should observe that $L^S > L_A^{MZ}$. Instead, if a specific characteristic is actually limiting the inequality gap, then we should have $L^S < L_A^{MZ}$.

These exercises are obviously hypothetical, because they do not account for all the possible changes in income (also within the groups) that would be associated with changes in the population structure. Furthermore, they are univariate, as they do not consider the correlation between the different dimensions. The fraction of inequality explained by one of them could therefore capture other related characteristics as well.

Demographic characteristics seem to play a minor role in explaining the higher inequality among southern residents (Figure 5; for more details about the groups for each dimension see Table A1 in the Appendix).[10] The differences across areas with respect to a detailed classification of household types (accounting for the combination of size, structure, age and gender of the reference person) do not seem, overall, to explain inequality differences between the two areas. The presence of larger households, which not only have lower income but are also more unequal between themselves, contributes positively to the higher income dispersion (as $L^S > L_A^{MZ}$), but the impact is quite small. Differences across areas in the reference person's age and country of birth instead provide a negative contribution to inequality in the South. Reference persons in southern regions are older and more likely to be born in the country, and both characteristics are associated with higher income. The impact on inequality is nevertheless small, in particular for the differences in country of birth. This is due to the fact that the income gap between foreign- and native-born residents is smaller in the South (Table A1 in the Appendix). While in the Centre-North the employment rate of adult household members (for households with at least one member aged 18-64) is lower if the reference person was born abroad, the opposite is true in southern regions.[11] Furthermore, while in the Centre-North, the inequality within the group of households with a foreign-born reference person is larger than that within the group with a native-born person, in the South the dispersion is larger in the latter group. Panel (a) of Figure 6 shows that the combined contribution (to southern inequality) of

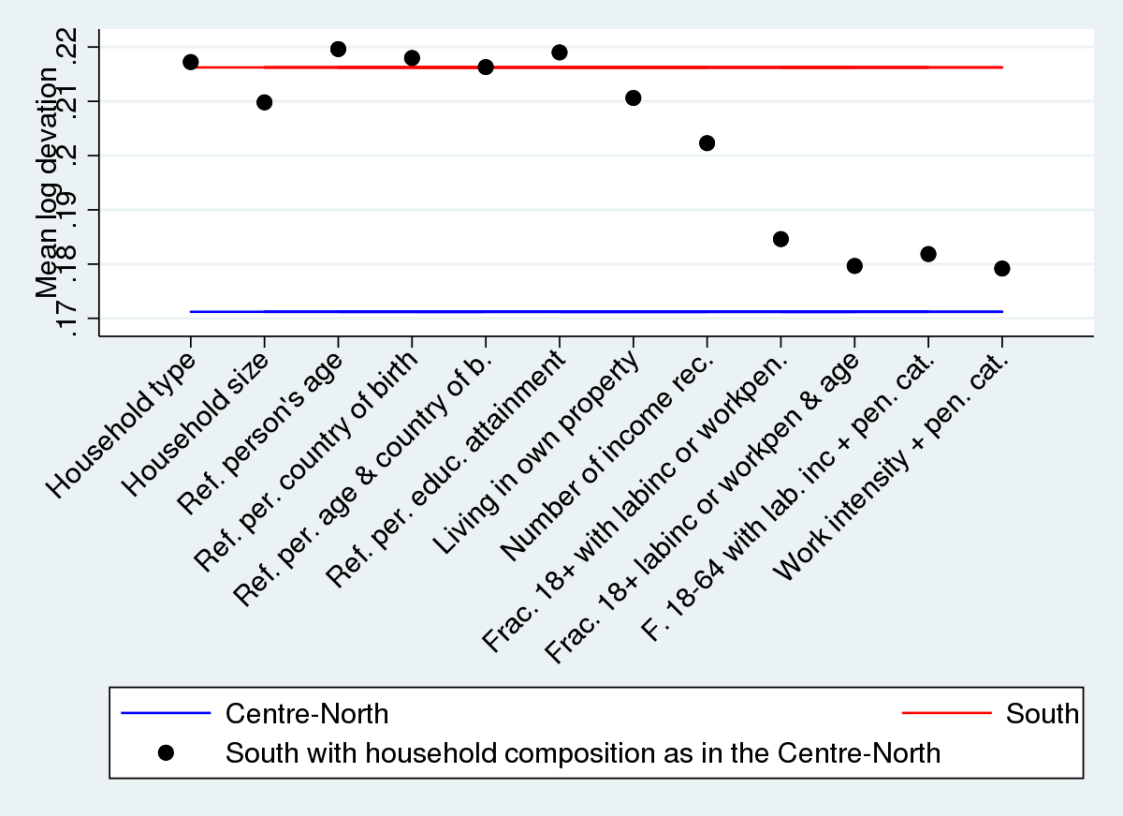
[10] Housing wealth is also not relevant, because the difference in the fraction of individuals living in owned property is similar in the two areas.

[11] As in the rest of the paper, averages are calculated using the person as a reference (even if the variable is defined at household level). The household-level average of the employment rate is similar for households with a foreign- or native-born reference person, while in the South it is still larger for those with a foreign-born reference person.

the reference persons' age and country of birth has been small over the entire period, even if both characteristics are important in explaining the evolution of the overall inequality at the national level.

Differences in the reference persons' educational attainment provide a negative contribution to inequality in the South. The income premium associated with the reference person holding a university degree is greater in the South (Table A1 in the Appendix). Therefore, the smaller share of the population with tertiary education contributes to limiting overall inequality.

Figure 5: Inequality in the South and household characteristics, 2016

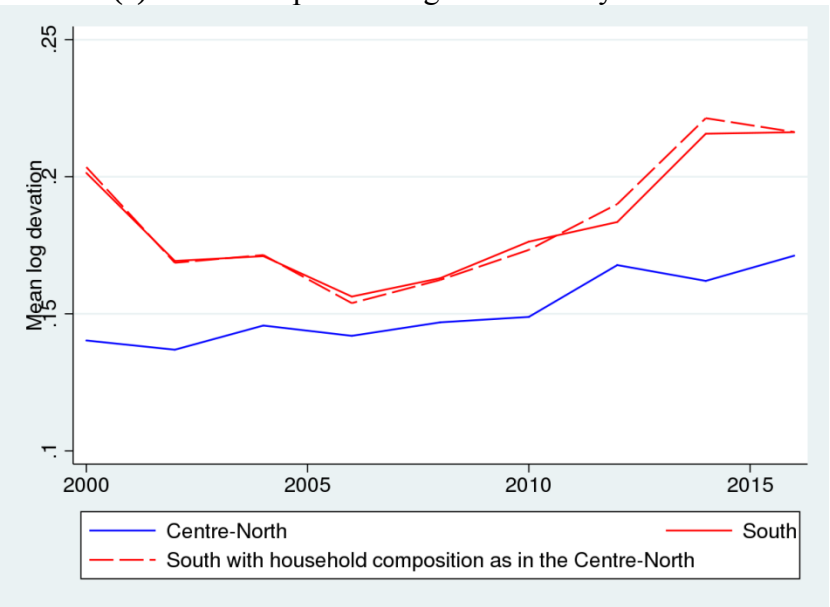


Note: The mean logarithmic deviation in the South with household composition as in the Centre-North is calculated using eq. (3). 'labinc' stands for labour income; 'work pen' for work-based pension; 'pen. cat.' indicates that households with a recipient of a work-based pension are considered as a separate category. For the details of the groups in which each dimension has been split, see Table A1 in the Appendix.

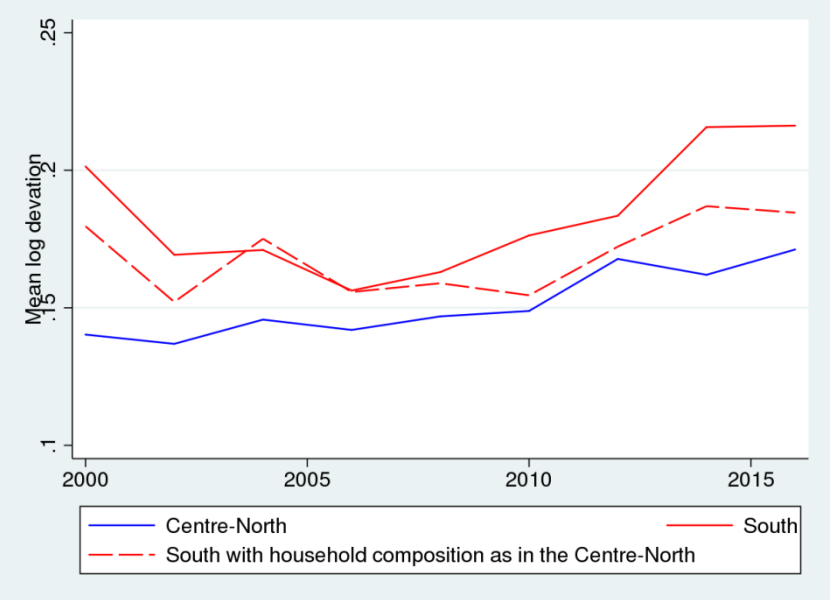
Variables relating to the receipt of labour income are more relevant in explaining the higher income inequality. As highlighted in Section 3.1, geographical differences are particularly marked in the fraction of adult household members that earn labour income or are recipients of a work-based pension inequality. By changing the population shares of southern households to mimic the distribution of this characteristic as observed in the Centre-North, the mean logarithmic deviation in the South shrinks by 15 per cent. Results looking at other dimensions of labour and pension income lead to similar conclusions, even when we consider the measure of work intensity based on the number of months spent in employment. Panel (b) of Figure 6 shows that the relevance of employment in explaining the higher inequality among southern households grew from the crisis up to 2016.

Figure 6: Inequality in the South with households' characteristics as in the Centre and North, 2000-2016

(a) Reference person's age and country of birth



(b) Fraction of adult recipients of labour income or a work-related pension



Focusing on the exercise where we change the share of adult household members that earn labour income or are recipients of a work-related pension, it is possible to calculate by how much national inequality would change, keeping the Centre-North contribution fixed. The contraction in the within-South's contribution is strengthened by the reduction in the average income gap between the two areas (that would shrink from 39 to 27 per cent). As a consequence, overall inequality in Italy would decrease by 12 per cent.

3.3 The role of employment levels

The univariate decompositions highlight the role played by employment rates and, more in general, work intensity. However, these decompositions have some limitations. First of all, by changing the composition of the southern population along one dimension we also implicitly change their characteristics across other related dimensions, such as education. Secondly, we might want to understand what would happen if, instead of changing employment levels, we applied the wage structure observed in the Centre-North of the country. Finally, spatial equilibrium models for the labour market suggest that one way to achieve higher employment rates is to provide for a contraction in earnings in the South (see Accetturo et al., 2009; Boeri et al., 2019), and therefore it is important to simulate what would happen if we decreased hourly wages when we increase work intensity.

We try to address these limitations with a static microsimulation in which we change southern households' work hours and their average hourly wage to the level of the Centre-North, while keeping other socio-demographic characteristics unchanged. We only focus on the work hours of household members of working age (and their average hourly wage in the household). As we are interested in inequality, we need to apply the distribution of work hours and average hourly wage observed in the Centre-North to southern households, and not only to change their average values. For this purpose we employ a matching-based approach, based on a series of steps performed only for households with at least one person aged 18-64:[12]

1. To each household in area A we assign the average work hours across components aged 18-64 of a matched household in area B.
2. To each household in area A with positive work hours we assign the average hourly wage of a matched household in area B.[13]
3. We recalculate the labour income of 18-64 members by using matched work hours and hourly wages and we use it to redefine total household income (and then to make it equivalent using the OECD scale).[14]

Given that the matching procedure never provides a perfect reconstruction of the underlying distributions and includes a random component, we start from two simulations in which we re-impute values for each area. For the Centre-North we follow steps 1-3 by using the Centre-North for both A and B, and similarly for the South.

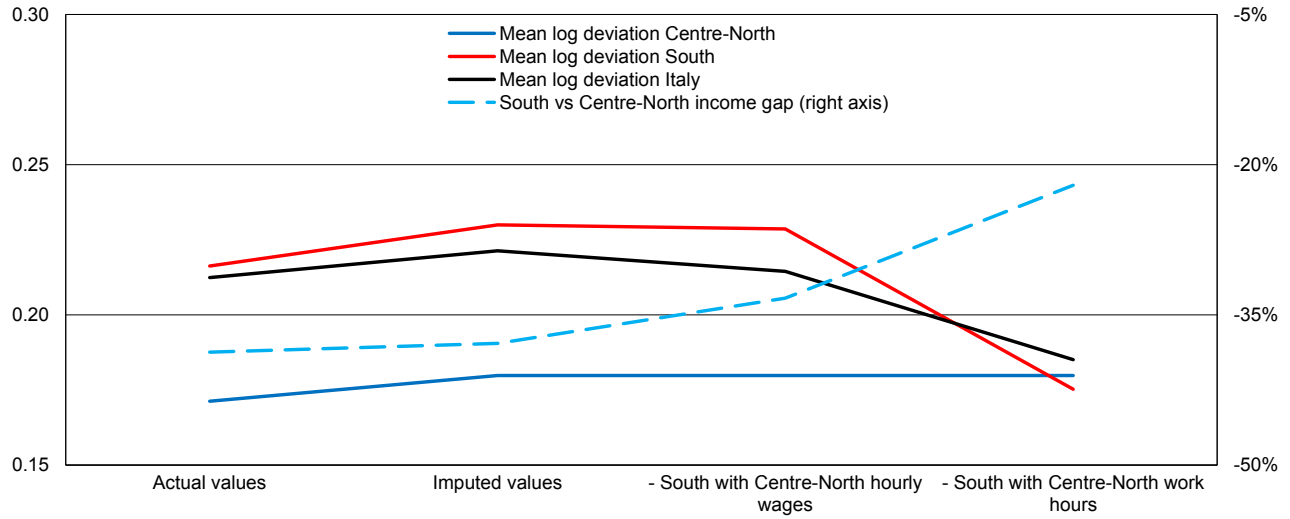
As we keep the socio-demographic composition of southern households fixed, when we impute to households in area A hours of work and hourly earnings from B, we want to use similar households along a vector X of characteristics (e.g. reference persons' education). In this way the simulation also imputes the association between hours of work (or hourly earnings) and X, as observed in area B (e.g. we impute the education premium for university graduates observed in area B). For this

[12] We work at the household level because this allows us to take into account the redistribution of income due to pooling resources. We also tried with individual-based simulations, but the imputed values tended to display a very high level of inequality with respect to the actual values.
[13] Hourly earnings are net of taxes.
[14] Given that the mean logarithmic deviation is strongly influenced by low values, we still need to apply bottom-coding. However, to avoid this bottom-coding changing across simulations, we use the 2nd percentile calculated as in the original sample (€1,500).

purpose, matches for step 1 are chosen by looking at the closest potential match in terms of the Mahalanobis distance calculated on *X*, which includes: reference person's age, country of birth and educational attainment (in dummies: primary school, middle school, high school, university); household size; number of household members aged 18-64; household transfer income; household pension income; household capital income; and a dummy for living in owned property.[15] In step 2, after having assigned hours of work to southern households, we include this variable among the covariates *X* used in the match (to account for the relationship between hourly earnings and work intensity). To avoid multiple households in area A being associated with the same household in B, the matches are run sequentially, household by household. Starting from the first household on the list for area A, the best closest match from B is chosen, but this household in B is then flagged as unavailable for any other match.[16] The order of households in the list is chosen randomly, and therefore we perform all simulations 100 times and take average values.

Figure 7 displays how inequality, as measured by mean logarithmic deviation (averaged across replications), would change in the different scenario. The simulation itself slightly increases inequality with respect to actual data. Imputed values are obtained by simulating hours of work and hourly earnings using households in the same area as matches (for instance, for the South we follow points 1-3 above setting A=South and B=South). Hence any comparison with the following policy scenarios should start from these simulated values.

Figure 7. Matching-based simulations, 2016



Note: all values are averages across 100 replications.

In the first policy scenario we impute to the South the hourly wage observed in the Centre and North (while work hours are still imputed in the South). As expected, the between-areas gap shrinks (by 5 percentage points). Inequality in the South is left almost unchanged. Overall, national inequality decreases by 3 per cent, from 0.221 to 0.214.

[15] The matches are chosen using the mahapick and mahaselectunique programs in Stata, written by Kantor (2006).
[16] This is not possible when we increase the employment level in the South to the levels observed in the Centre-North, while keeping the hourly wages of the South. This is because to impute hourly wages we need to get a match in the South, but the number of households with employed members in the actual data is lower than in the simulated one. For these simulations we therefore allow households in area B to be picked up twice (still randomly).

In the second policy scenario we impute to the South the hours of work observed in the Centre-North, while we impute hourly wages from the South itself. These changes have a strong impact on both between- and within-area inequality. With respect to the (simulated) status quo, the income gap shrinks by 16 percentage points, far more than in the previous scenario. The reduction in the number of households with low work intensity brings about a drop in the dispersion of working hours within the South, which leads to a large decrease in the mean logarithmic deviation (by 24 per cent, from 0.230 to 0.175). Overall, inequality at the national level drops by 16 per cent.

The matching-based simulations focus only on changes in labour income, leaving other components such as transfers and capital income untouched. To account for the fact that other income components may change when we change working hours, the focus should be on the distributional relationship between the overall household equivalent income *y* and the vector of characteristics *Z*=<*h*,*X*>, where *h* are per-capita hours of work by household members aged 18-64 and *X* the vector of socio-demographic characteristics (as before). In our case, we would like to estimate how the distribution of *y* across southern households changes when the distribution of *h* is changed to reflect the one observed in the Centre-North, while keeping *X* fixed.

To do so we follow a reweighting procedure as described in Fortin and Lemieux (2011, pp. 85-86). Formally, we consider two areas A and B, and let F_{v_a} be the distribution of variable *v* in area *a*. Our object of interest is the counterfactual distribution of *y* in area *A* if the distribution of *h* was as in area *B*, but conditional on leaving *X* unchanged:

$$F_{Y_A^{C,h}}(y) = \int F_{Y_A|Z_A}(y|Z)dF_{Z_B}(h|X)dF_{X_A}(X) = \int F_{Y_A|Z_A}(y|Z)\Psi_{h|X}(X)F_{X_A}(Z) \qquad (4)$$

where $\Psi_{h|X}(X)$ is a reweighting factor that can be written as

$$\Psi_{h|X}(X) \equiv \frac{F_{X_B}(h|X)}{F_{X_A}(h|X)} \qquad (5)$$

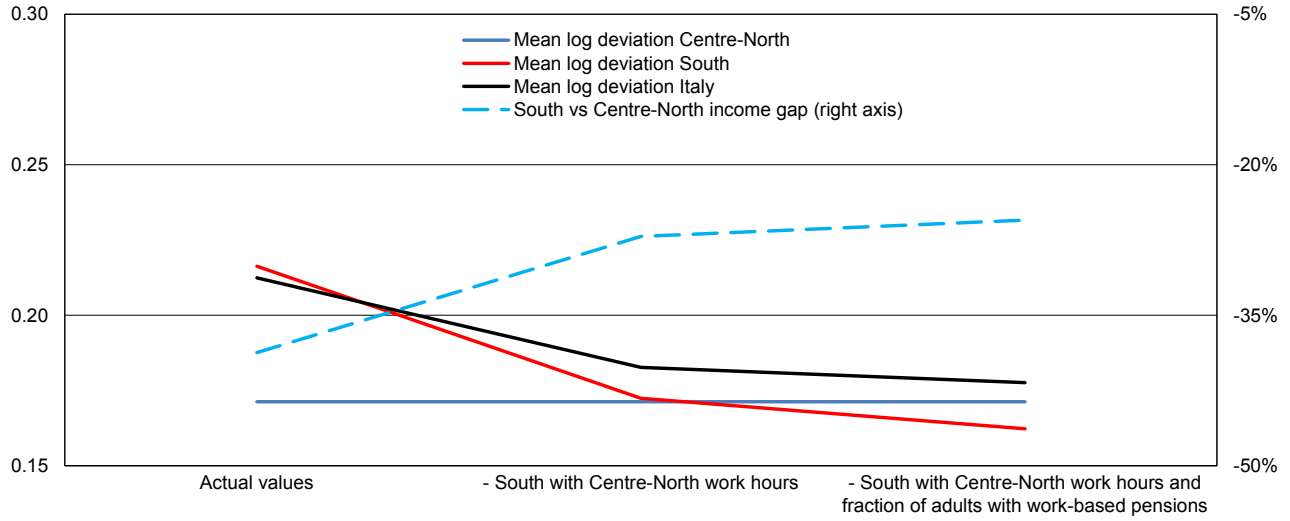
If we know $\Psi_{h|X}(X)$, then we can just use it as a sample weight for southern households and calculate all the statistics of interest.[17] The authors show that $\Psi_{h|X}(X)$ is equal to ratio of two reweighting factors, one based on predicting the probability of being in area B on the basis of the entire vector of variables *Z*=<*h*,*X*>, and the other predicting the same probability but using *X* alone. We estimate both models using a probit model in which we consider: reference person's age, country of birth and educational attainment (in dummies: primary school, middle school, high school, university); household size; number of household members aged 18-64; a dummy for living in own property; and the fraction of adults with a work-based pension.[18] Table A.3 in the Appendix

[17] Given that the sample is stratified, the reweighting factor is multiplied by the original sample weights to account for this.

[18] We do not consider household transfer income, household pension income and household capital income because in this exercise we do not want to keep them fixed at the pre-existing level. We also include a set of interactions between work hours and the reference person's educational attainment, age, and number of household members aged 18-64; between the fraction/share of adults with a work-based pension and work hours, the reference person's educational attainment, and number of household members aged 18-64; between a dummy for the reference person being a university graduate and a dummy for him/her being native-born. We also include a triple interaction between a dummy for the reference person being a university graduate, a dummy for him/her being foreign-born and work hours. These interactions are progressively included until the re-weighted sample average of the variables in h in area A become

shows evidence that the reweighting procedure works, by showing that the averages of the variables in *X* do not change significantly, while *h* is brought to the levels observed in the Centre-North. [19]

Figure 8. Reweighting results, 2016



The results (Figure 8) for the change in work hours are very close to those obtained by matching, with a drop in inequality at the national level of 14 per cent, and therefore confirm our main analysis.

As discussed in the introduction, one policy that could potentially boost employment in the South would be to reduce the bite of collective bargaining, thereby allowing for a contraction of local labour cost in regions with lower productivity. We abstain here from formulating hypothesis about the size of the labour cost cut that would allow hours of work to reach the levels of the Centre-North. In fact, the required increase in hours would be around 49 per cent, hence a policy relying only on wage adjustments would need to assume extreme changes in wages or implausibly large (unconditional) labour demand elasticities. Nevertheless, reducing relative wages could be part of a larger package of policies aimed at boosting employment in the South. We therefore try to assess the impact of cuts of different size on overall inequality[20].

---

close (and not statistically different) to the ones in area B, while the other variables are kept to similar levels as before. When we estimate the reweighting factor without *h*, all interactions including it are removed.

[19] In the Appendix Figure A1 we also show that the predicted probability of being in the Centre-North (based on the entire vector of variables Z=<h,X>) spans from small to high values in both areas and therefore there is common support. One concern is that the reweighting factor $\Psi_{h|X}(X)$ is quite large in some cases, because the two predicted probabilities (with and without h, see eq. 5) are rather different. In Figure A2 we show that the overall distribution of sample weights is similar to the distribution of reweighted sample weights. However, there are some households where the reweighted sample weight is quite large. In Appendix Figure A3 we show that results are similar if we censor the reweighted sample weight at the 99th percentile (which is 21.8; as a comparison, the 99th percentile of the original sample weight is 17.6), or we censor the weights below the 1st percentile (by replacing them with this value) and above the 99th percentiles of the original sample weight distribution.

[20] It should be taken into account that the exercise considers a reduction in net wages, but the relevant variable for labor demand is total labour cost. Given that the income tax is progressive, the proportional cut in gross earnings is larger than the associated cut in net wages. Furthermore, a comprehensive strategy to rise employment could accompany a reduction in local contractual gross wages with fiscal measures aimed at reducing the impact on low-wage workers' take-home pay. Such a strategy might be necessary to promote labour supply, in particular for low skilled women that are much less likely to be employed in the South (see Colonna and Marcassa, 2015, for a discussion of the role of the Italian tax system in discourage unskilled women's labour supply).

In Figure 9 we simulate an additional scenario in which we still simulate (using the reweighting exercise) an increase in southern hours of work to the levels of the Centre-North, but we cut southern hourly earnings by 20 per cent (approximately the difference in labour productivity).[21] This further reduces inequality within the South, as labour income (which is the component contributing most to inequality) drops. This contraction brings the geographical average gap back to 36 per cent, therefore increasing the between-area inequality. Nevertheless, the wage cut only partially erodes the beneficial effect on national inequality of an increase in work intensity among southern households (the drop in national inequality is 10 per cent). If we cut average earnings by 30 per cent, the between-area component increases further, but national inequality is still lower than in the status quo. Only with a cut of 40 per cent would we get back to the initial inequality levels.

Figure 9. Reweighting results, with wage cuts, 2016

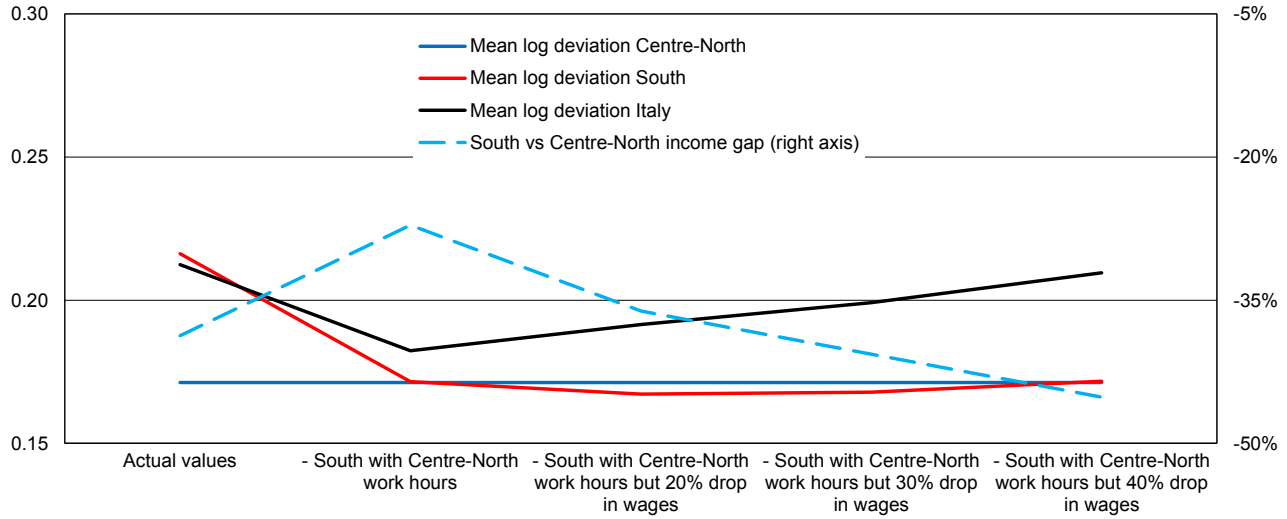
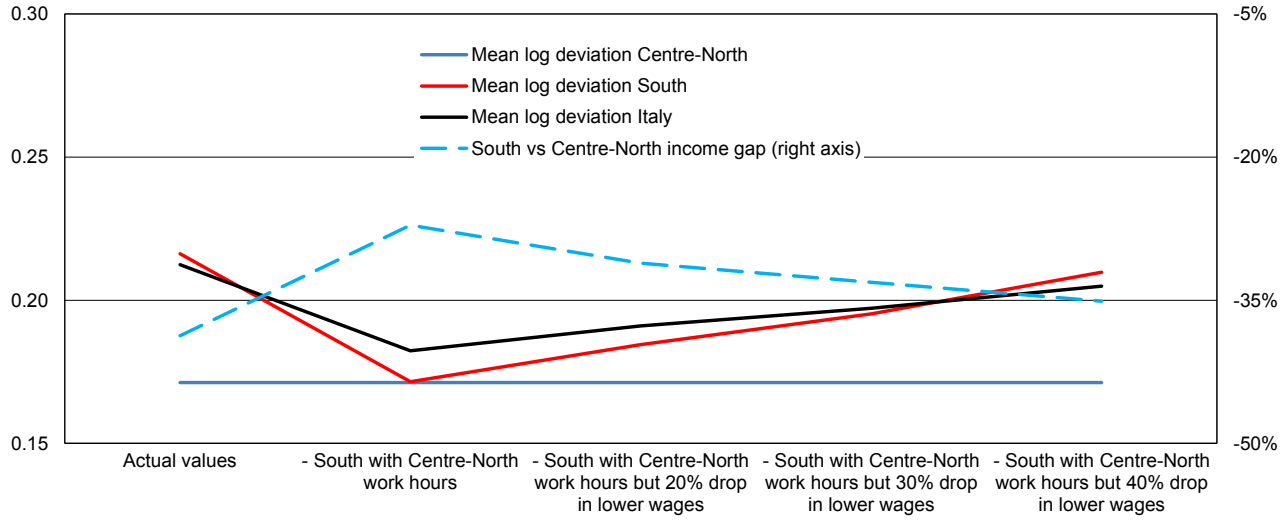


Figure 10. Reweighting results, with wage cuts only for lower wages, 2016



A proportional cut in hourly wages may be unlikely. Wage bargaining might be more binding for lower hourly wages. In Figure 10 we impose cuts only on those households whose average hourly earnings (across household members) is below the median (calculated within the South). In this

[21] According to the latest available data from the National Institute of Statistics (Istat), in 2017 value added per worker was 23 per cent lower in the South (25 per cent of the value added per hour of work, year 2016).

case, a wage cut increases not only between-area inequality, but also the within-area inequality component, if compared with a scenario in which the rise in employment is attained without any reduction in average labour earnings. Nevertheless, in this case too, a sizeable drop is necessary to offset the beneficial effect on income inequality of rising employment levels.

One final issue is that an improvement in labour market conditions in the South will lead, in the long run, to a larger share of elderly people entitled to a work-based pension. The reweighting approach allows us to assess what would happen by changing both the hours of work of working age household members (h) and the share of adults receiving such pensions, while keeping other characteristics in X unchanged. The results are reported as for the last scenario in Figure 8. The impact is not too strong with respect to the simulation in which we changed only hours of work. Inequality within the South shrinks further (the mean logarithmic deviation drops from 0.172 to 0.162) and the gap decreases by one percentage point. The limited impact of these changes is partially due to the fact that, during the last twenty years, inequality among the elderly (as well as poverty) has been decreasing, thanks to repeated rises in minimum and social pensions.

4. Conclusions

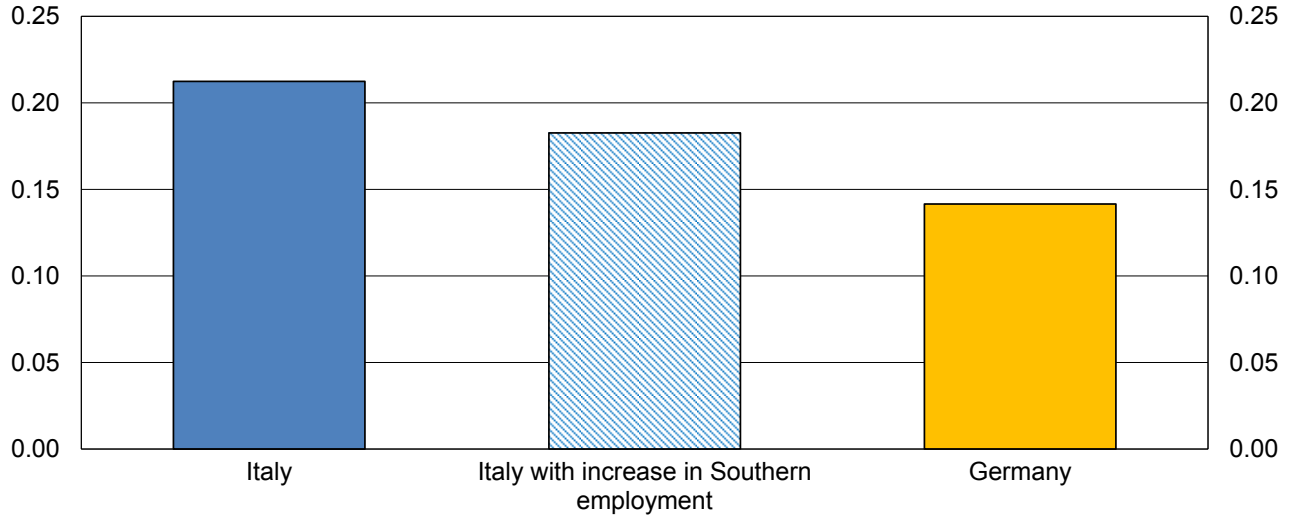
In this paper we reassessed the role of geographical imbalances in explaining Italy's high household income inequality. Recent trends have led to an increase in within-area inequality, which is structurally higher among southern households. The double-dip recession hit the bottom of the income distribution more severely, which experienced a strong contraction in both labour and transfer income, especially in the South. The North-South average income gap, despite having been relatively stable since the early 2000s, still accounts for more than one tenth of the overall inequality.

We contributed to the literature by providing a detailed analysis of the role of the low employment intensity among southern households in explaining national inequality. Starting from univariate decompositions, we showed that the fraction of adult members in employment – which is also reflected in the receipt of work-related pensions – is key in accounting for the higher inequality in the South and for the large geographical divide.

By means of simulations based on matching and reweighting we estimated that, while keeping other characteristics (such as demographics and educational level) constant, a shift in the distribution of work hours among southern households towards that observed in the Centre-North would lead to a large reduction in Italian inequality of approximately 15 per cent. As a result, the inequality gap between Italy and Germany (according to the latest available estimates by Centro Studi Confindustria, 2015) would be approximately halved (see Figure 11). If wages in the South decreased by as much as 20 per cent to facilitate this shift, the contraction in the geographical divide would be smaller, but the overall reduction in national inequality would still be sizeable (around 10 per cent). In order for the effect of the increase in hours of work to be completely offset by a wage reduction, the cut should reach 40 per cent. A similar result arises if the cut is concentrated among households with lower hourly earnings, although in this case the reduction in wages also offsets the contraction in inequality within the South, and not only the decrease in the North-South average income differential.

These results provide useful insights for designing policies aimed at promoting a more equitable income distribution. Although they by no means imply that a reform of redistributive policies is unnecessary, they point out that such a reform should be designed in order to minimize the risk of discouraging employment. Furthermore, our results suggest that policies aimed at raising the employment rate in Southern regions are key to curb market income inequality in Italy. We showed that substantial gains would be achieved even in the case this required a significant reduction of local relative wages.

Figure 11. Mean logarithmic deviation in Italy and Germany



Note: own calculations on SHIW 2016 for Italy; for Germany, we calculate the index using the fraction of Germany vs Italian mean logarithmic deviation calculated by Centro Studi Confindustria (2015).

References

Accetturo, A., Dalmazzo, A., De Blasio G., and R. Torrini (2009), North-South imbalances: a spatial general equilibrium approach, in Banca d'Italia (2010), Il Mezzogiorno e la politica economica in Italia, Seminari e convegni, n. 4.

Acciari, P. and S. Mocetti (2013) Una mappa della disuguaglianza del reddito in Italia, Questioni di Economia e Finanza (Occasional Papers), Banca d'Italia, 208

Baldini, M., Bosi, P. and S. Toso (2002) - Targeting Welfare in Italy: old problems and perspectives on reform, Fiscal Studies, 23: 51-75

Baldini, M. and S. Toso (2009) Diseguaglianza, povertà e politiche pubbliche, Bologna, Il Mulino.

Banca d'Italia (2010) Il Mezzogiorno e la politica economica dell'Italia, Seminari e convegni, n°..

Bodo, G. and P. Sestito (1991) Le vie dello sviluppo, il Mulino, Bologna.

Boeri, T., Ichino, A., Moretti E. and J. Posch (2019) Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces, NBER Working Paper No. 25612.

Brandolini, A. (2009) La disuguaglianza dei redditi personali: perché l'Italia somiglia più agli Stati Uniti che alla Germania? in R. Catanzaro e G. Sciortino (a cura di), La fatica di cambiare. Rapporto sulla società italiana, pp. 133-153, Bologna, Il Mulino.

Brandolini, A. and R. Torrini (2010) Disuguaglianza dei redditi e divari territoriali: l'eccezionalità del caso italiano, la Rivista delle Politiche Sociali, vol. 3/2010

Brandolini, A. and G. D'Alessio (2001) Household Structure and Income Inequality, LIS Working papers 254, LIS Cross-National Data Center in Luxembourg.

Centro Studi Confindustria (2015) Distribuire meglio, crescere di più, Scenari economici n. 23.

Ciani, E., David, F., and G de Blasio (2019) Local responses to labor demand shocks: A Re-assessment of the case of Italy, Regional Science and Urban Economics, forthcoming.

Colonna F. and S. Marcassa (2015) Taxation and female labor supply in Italy, IZA Journal of Labor Policy, 4(1): 1-29.

European Commission (2014) Guidelines on Regional state aid for 2014-2020, Competition policy brief, Issue 14.

Faini, R. (1999), Trade unions and regional development, European Economic Review, 43: 457-474.

Fortin, N. Lemieux, T., and S. Firpo (2011) Decomposition Methods in Economics, in O. Ashenfelter and D. Card (Eds.), Handbook of Labor Economics, vol. 4A, pp 1-102. Amsterdam: North-Holland

Gorrieri, E. (1972) La giungla retributive, Bologna, Il Mulino.

Kantor, D. (2006) MAHAPICK: Stata module to select matching observations based on a Mahalanobis distance measure, Statistical Software Components S456703, Boston College Department of Economics.

Peragine, V. (2017) La disuguaglianza in Italia: le tendenze nascoste dagli indici sintetici, url: https://www.eticaeconomia.it/la-disuguaglianza-in-italia-le-tendenze-nascoste-dagli-indici-sintetici

Appendix

Table A1. Distribution of equivalent disposable household income across individuals living in the Centre-North and the South, by socio-demographic characteristics; year 2016, per cent

	Fraction of the population		Relative wage compared to the area mean	
	Centre-North	South	Centre-North	South
Household type				
Male, single, <=65 years	4.8	2.8	101.9	115.5
Female, single, <=65 years	4.2	2.4	97.7	101.3
Male, single, >65 years	1.9	1.3	111.8	136.2
Female, single, >65 years	5.0	4.7	92.9	107.0
Couple, no children, reference person <=65	7.8	5.9	126.4	124.6
Couple, no children, reference person >65	10.8	8.5	117.7	121.1
Couple with one child	16.8	18.0	107.6	111.4
Couple with two children	23.4	27.1	96.8	93.8
Couple with three or more children	9.4	10.3	76.5	68.8
Single parent with a child aged <18	1.3	0.3	88.9	76.9
Single parent aged <=65 with a child>=18	1.6	1.6	90.5	75.6
Single parent aged >65 with a child>=18	0.9	1.4	88.3	95.7
Single parent with more than one child	3.3	3.8	64.0	69.2
Other types	8.7	11.7	93.5	100.5
Household size				
1	15.9	11.3	99.1	111.2
2	24.3	19.5	115.8	116.2
3	22.3	22.4	102.3	108.0
4	24.9	31.5	95.9	95.9
4+	12.7	15.4	74.7	68.1
Reference person's age				
Up to 30	3.7	4.2	82.0	84.4
31-40	14.6	14.0	85.7	84.6
41-50	27.1	23.1	97.3	94.3
51-65	30.5	33.6	106.9	103.7
>65	24.2	25.1	105.8	111.5
Reference person's country of birth				
Italy	87.3	96.0	106.5	101.5
Other countries	12.7	4.0	55.0	63.1
Reference person's age and country of birth				
<=44, Italy	7.0	2.4	53.1	58.0
<=44, other countries	20.9	25.4	100.5	92.4
[45,64], Italy	5.3	1.4	54.2	69.2
[45,64], other countries	41.3	43.8	109.4	100.2
65+	25.6	27.0	106.6	112.1

(continues below)

Reference person's educational attainment				
Less than secondary	54.4	63.7	80.4	78.2
Secondary	29.6	26.4	106.9	120.8
University	16.1	9.9	153.9	184.9
Living in own property				
No	28.6	32.5	67.4	72.8
Yes	71.4	67.5	113.0	113.1
Fraction of adult members recipients of labour income or work-based pension				
0	5.9	14.0	50.1	50.0
[0,50) per cent	27.6	49.2	77.5	84.4
[50,100] per cent	66.5	36.8	113.8	139.9
Fraction of adult members recipients of labour income or work-based pension, combined with reference person's age				
0, up to 65	2.8	8.6	31.8	31.6
Between 0 and 50, up to 65	21.8	38.6	76.0	80.7
Between 50 and 100, up to 65	49.8	25.8	110.9	139.1
0, over 65	3.1	5.4	66.4	79.2
Between 0 and 50, over 65	5.8	10.6	83.1	97.8
Between 50 and 100, over 65	16.7	11.0	122.2	142.0
Fraction of working-age members recipients of labour income, combined with a separate category for households with no working-age member but at least one recipient of a work-based pension				
0	6.1	14.8	51.6	53.2
[0,50) per cent	25.0	40.5	82.0	83.8
[50,100] per cent	49.5	27.8	116.9	139.5
0 but with recipients of a work-based pension	19.4	16.9	108.6	114.8
Work intensity (1)				
Low	6.8	17.7	49.4	51.8
Medium	25.1	39.3	83.4	86.5
High	48.1	25.4	112.7	145.4
Low but with recipients of a work-based pension	20.0	17.7	107.5	113.0

Note: (1) Work intensity is calculated as the total number of months worked by members aged 18-64 divided by the total number of months available to them (12 months times their number). High refers to a value above 0.5; medium to [0.2,0.5]; low to a value below 0.2. Households without working age members are in the low group, but all those cases with a low work intensity and a recipient of a work-based pension are in a separate category.

Table A2. Decomposition of within-area mean logarithmic deviation by socio-demographic characteristics, 2016

Inequality component	Centre-North		South	
	Value	Share per cent	Value	Share per cent
Household type				
Intra-group	0.160157	93.5	0.200918	92.9
Inter-group	0.011055	6.5	0.015329	7.1
Household size				
Intra-group	0.163267	95.4	0.202324	93.6
Inter-group	0.007946	4.6	0.013923	6.4
Reference person's age				
Intra-group	0.167697	97.9	0.21168	97.9
Inter-group	0.003515	2.1	0.004567	2.1
Reference person's country of birth				
Intra-group	0.150666	88.0	0.212507	98.3
Inter-group	0.020547	12.0	0.00374	1.7
Reference person's age and country of birth				
Intra-group	0.149494	87.3	0.209836	97.0
Inter-group	0.021719	12.7	0.006411	3.0
Reference person's educational attainment				
Intra-group	0.14114	82.4	0.170262	78.7
Inter-group	0.030073	17.6	0.045985	21.3
Living in own property				
Intra-group	0.146012	85.3	0.196224	90.7
Inter-group	0.025201	14.7	0.020023	9.3
Fraction of adult members recipients of labour income or work-based pension				
Intra-groups	0.145826	85.2	0.159055	73.6
Inter-groups	0.025387	14.8	0.057192	26.4
Fraction of adult members recipients of labour income or work-based pension, combined with reference person's age				
Intra-group	0.141159	82.4	0.142901	66.1
Inter-group	0.030054	17.6	0.073347	33.9
Fraction of working-age members recipients of labour income, combined with a separate category for households with no working-age member but at least one recipient of a work-based pension				
Intra-group	0.151954	88.8	0.167139	77.3
Inter-group	0.019258	11.2	0.049108	22.7
Work intensity				
Intra-group	0.1495823	87.4	0.1597211	73.9
Inter-group	0.02163	12.6	0.056526	26.1

Table A3. Reweighting exercise, 2016

MZ	Reference person's age	Household size	1[living in own property]	1[no members aged 18-64]	1[one member aged 18-64]	1[two members aged 18-64]	1[three members aged 18-64]	1[[four + members aged 18-64]	1[Reference's person no degree]	1[Reference person's primary school degree]	1[Reference person's middle school degree]	1[Reference person's high school degree]	1[Reference person's high university degree]	1[Reference person born abroad]	Fraction of adults with work-based pensions	Total hours of work of members aged 18-64
Panel A: Averages using actual sample weights																
Centre-North	54.5	3.0	0.29	0.17	0.19	0.39	0.15	0.11	0.01	0.12	0.41	0.30	0.16	0.13	0.21	1104
South	55.2	3.3	0.32	0.14	0.15	0.37	0.17	0.17	0.04	0.18	0.42	0.26	0.10	0.04	0.14	751
Difference	0.8	0.3	0.04	-0.03	-0.04	-0.02	0.03	0.07	0.03	0.06	0.00	-0.03	-0.06	-0.09	-0.07	-353
p-value of difference	0.204	0.000	0.049	0.001	0.000	0.346	0.093	0.000	0.000	0.000	0.907	0.104	0.000	0.000	0.000	0.000
Panel B: Averages using actual reweighting weights for the scenario with adjusted work hours																
Centre-North	54.5	3.0	0.29	0.17	0.19	0.39	0.15	0.11	0.01	0.12	0.41	0.30	0.16	0.13	0.21	1104
South	55.6	3.2	0.31	0.15	0.15	0.36	0.17	0.18	0.04	0.17	0.39	0.28	0.11	0.05	0.15	1101
Difference	1.1	0.2	0.03	-0.02	-0.04	-0.03	0.02	0.07	0.03	0.05	-0.02	-0.02	-0.05	-0.08	-0.06	-3
p-value of difference	0.084	0.007	0.262	0.053	0.007	0.304	0.259	0.038	0.000	0.000	0.440	0.597	0.006	0.000	0.000	0.960
Difference in the difference between Panels B and A																
value	0.4	0.0	-0.01	0.01	0.00	-0.01	-0.01	0.00	0.00	-0.01	-0.02	0.02	0.02	0.01	0.01	351
per cent	0.6	-0.6	-3.4	7.0	2.2	-2.2	-3.9	1.1	-0.2	-3.5	-5.8	5.8	15.5	20.0	3.5	46.7
Panel B: Averages using actual reweighting weights for the scenario with adjusted work hours and fraction of adults with work-based pensions																
Centre-North	54.5	3.0	1.29	0.17	0.19	0.39	0.15	0.11	0.01	0.12	0.41	0.30	0.16	0.13	0.21	1104
South	55.5	3.3	1.32	0.15	0.15	0.36	0.17	0.18	0.04	0.17	0.39	0.29	0.11	0.04	0.19	1101
Difference	1.0	0.3	0.033	-0.024	-0.039	-0.031	0.018	0.075	0.033	0.046	-0.028	-0.004	-0.048	-0.084	-0.025	-3
p-value of difference	0.121	0.009	0.244	0.053	0.009	0.276	0.342	0.043	0.000	0.001	0.351	0.909	0.004	0.000	0.059	0.958
Difference in the difference between Panels C and A																
value	0.3	0.0	-0.01	0.01	0.00	-0.01	-0.01	0.01	0.00	-0.01	-0.03	0.03	0.01	0.00	0.04	351
per cent	0.5	0.6	-2.0	6.3	2.3	-2.8	-5.1	4.1	1.9	-6.7	-7.3	10.5	13.7	7.1	31.0	46.7

Figure A1. Distribution of the predicted probability of being resident in the Centre-North, probit estimates

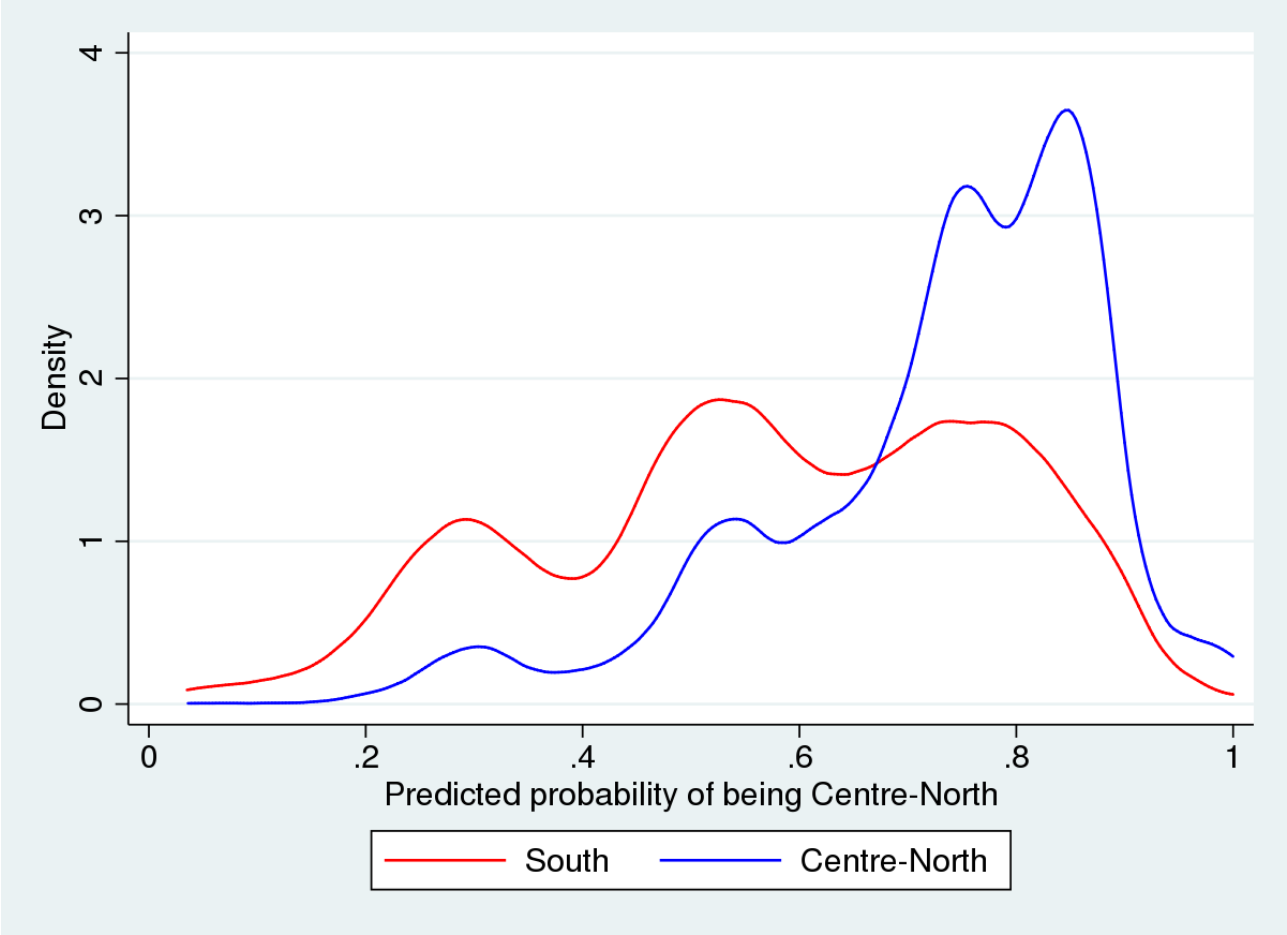


Figure A2. Distribution of the original sample weights vs reweighted sample weights

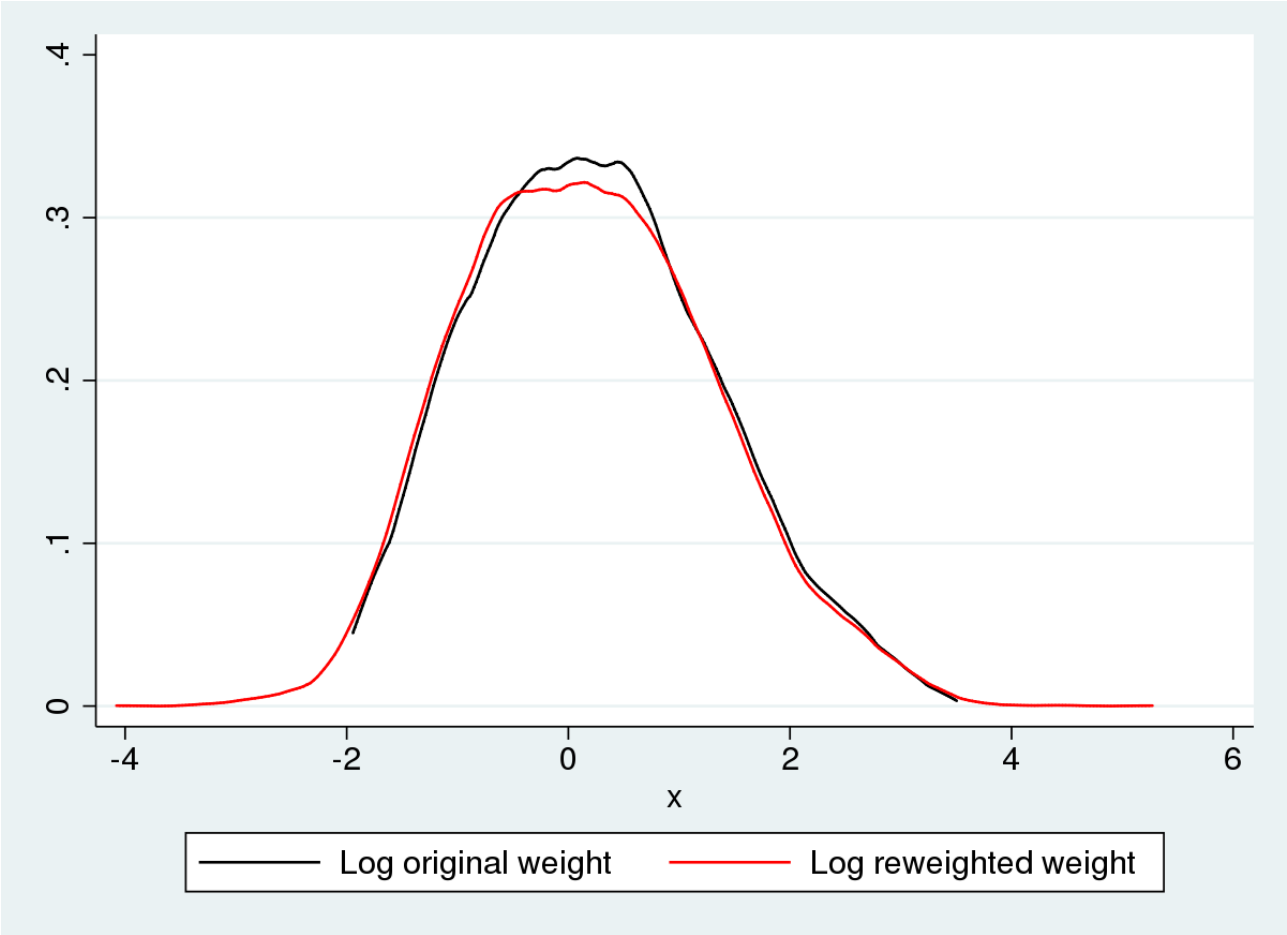
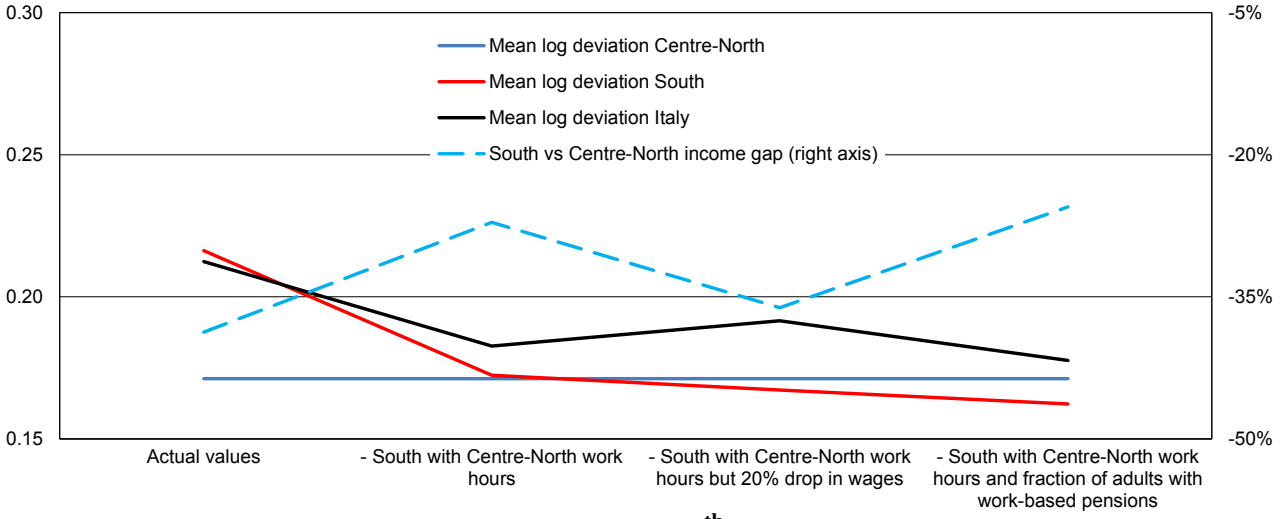
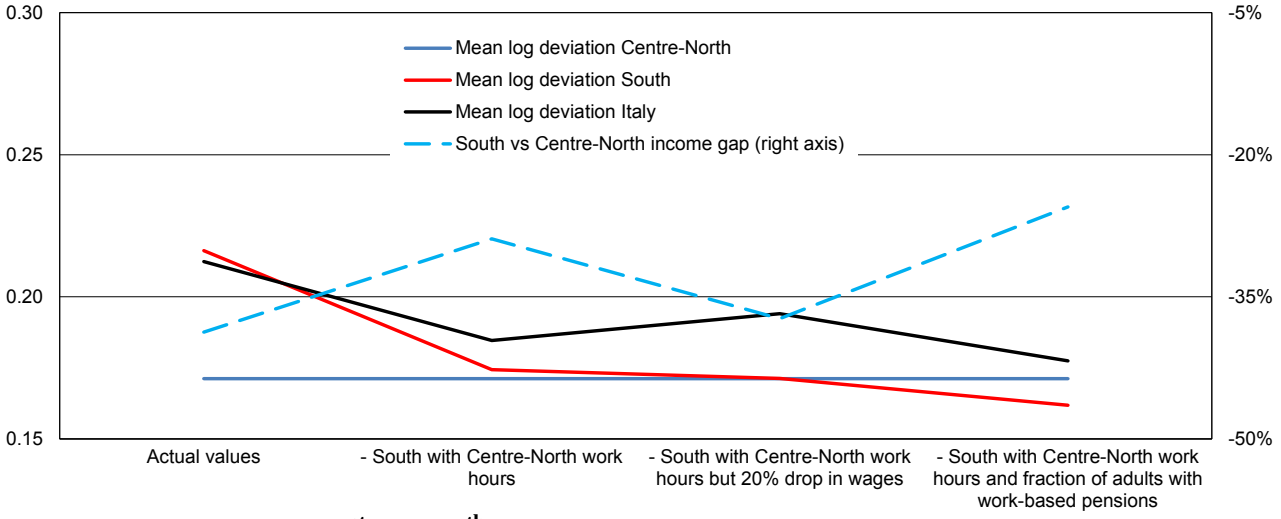


Figure A3. Reweighting results, with different treatment of the reweighted sample weights in the South, 2016

(a) no adjustment (as in the main text)



(b) censored at the 99th percentile



(c) censored at the 1st and 99th percentile of the original sample weight distribution