INPS

Istituto Nazionale Previdenza Sociale

giugno 2022 – numero 53

WorkINPS *Papers*

The Geographical Decomposition of Italian Inequality

Juraj Briskar

Edoardo Di Porto

Josè V. Rodriguez Mora

Cristina Tealdi

ISSN 2532 -8565

In copertina: uno storico "Punto cliente" a Tuscania
INPS, Direzione generale, Archivio storico

I WORKINPS PAPER

Le basi dati amministrative dell'*INPS* rappresentano una fonte statistica unica per studiare scientificamente temi cruciali per l'economia italiana, la società e la politica economica: non solo il mercato del lavoro e i sistemi di protezione sociale, ma anche i nodi strutturali che impediscono all'Italia di crescere in modo adeguato. All'interno dell'Istituto, questi temi vengono studiati sia dai funzionari impiegati in attività di ricerca, sia dai *VisitInps Scholars*, ricercatori italiani e stranieri selezionati in base al loro curriculum vitae e al progetto di ricerca presentato.

I **WORKINPS** hanno lo scopo di diffondere i risultati delle ricerche svolte all'interno dell'Istituto a un più ampio numero possibile di ricercatori, studenti e policy markers.

Questi saggi di ricerca rappresentano un prodotto di avanzamento intermedio rispetto alla pubblicazione scientifica finale, un processo che nelle scienze sociali può chiedere anche diversi anni. Il processo di pubblicazione scientifica finale sarà gestito dai singoli autori.

Maurizio Franzini

The Geographical Decomposition of Italian Inequality

Juraj Briskar

(London School of Economics and Political Science (LSE), London, WC2A 3PH, United Kingdom)

Edoardo Di Porto

(INPS, UCFS, Uppsala University, Universita' Federico II Napoli and CSEF.)

Josè V. Rodriguez Mora

(University of Edinburgh. 31 Buccleuch Place, Edinburgh)

Cristina Tealdi

(Heriot-Watt University, Edinburgh (UK) and IZA Institute of Labor)

INPS.0072.13/06/2022.0000485 E

The Geographical Decomposition of Italian Inequality*

Juraj Briskar† Edoardo Di Porto‡ José V. Rodríguez Mora§

Cristina Tealdi¶

June 10, 2022

Abstract

We show that geography plays only a marginal role in the determination of inequality between Italians, both in cross-section and in lifetime sense. Using social security data, we demonstrate that in spite of very large differences in average income between provinces, less than 4% of total cross-sectional inequality can be attributed to differences between provinces. When we calculate the lifetime income of a cohort of Italians (born in 1960), the share of variance explained by differences between provinces is 3.4% for the whole cohort and only 1.8% for males. For females, the number is substantially larger (10.2%). Thus, geography is a vector explaining inequality between Italians only in the sense that it affects female labor force participation. Finally, we also show that not only geography does not help predict a person's income, but the opposite is also

*We thank participants at the 2021 VisitINPS workshop, the VisitINPS seminar, and the 2022 MacCaLM conference for their useful comments. All remaining errors are our own. The findings and conclusions expressed are solely those of the authors and do not represent the views of INPS. La realizzazione del presente articolo è stata possibile grazie alle sponsorizzazioni e le erogazioni liberali a favore del programma "VisitINPS Scholars".

†London School of Economics and Political Science (LSE), 32 Lincoln's Inn Fields, Holborn, London, WC2A 3PH, United Kingdom.

‡INPS, UCFS, Uppsala University, Universita' Federico II Napoli and CSEF. The views expressed in this article are those of the author and do not necessary reflect the view of INPS.

§University of Edinburgh. 31 Buccleuch Place, Edinburgh. sevimora@gmail.com.

¶Heriot-Watt University, EH14 4AS Edinburgh (UK) and IZA Institute of Labor.

true: knowing the income of a person does not help much in predicting her province of birth or residence.

Keywords: lifetime income, variance decomposition, regional inequality.

La decomposizione geografica della disuguaglianza di reddito in Italia

Abstract

In questo articolo mostriamo come le differenze tra aree geografiche in Italia contribuiscano marginalmente alla disuguaglianza di reddito, sia quando consideriamo i redditi annuali, sia quando consideriamo i redditi dell'intera vita lavorativa. Utilizzando dati forniti dall' Istituto Nazionale della Previdenza Sociale (INPS), dimostriamo come nonostrante ci sia una grande differenza di reddito tra province, meno del 4% della disuguaglianza totale dei redditi annuali puo' essere attribuita a differenze tra province. Quando utilizziamo i redditi dell'intera vita lavorativa per una coorte di individui nati nel 1960, la percentuale della varianza attribuibile alle differenze tra province e' del 3.4% per l'intera coorte, e del 1.8% per i soli uomini. Per le donne, la percentuale sale al 10.2%. Pertanto le differenze tra province spiegano la disuguaglianza di reddito in Italia solo nel senso che sono associate a grandi differenze nella partecipazione alla forza lavoro femminile. Infine, mostriamo che non solo le differenze geografiche non aiutano a spiegare le differenze di reddito, ma e' anche vero il contrario: conoscere il reddito di un individuo non aiuta a identificare la sua provincia di nascita o di residenza.

Keywords: reddito della vita lavorativa, decomposizione della varianza, disuguaglianza regionale.

1 Introduction

If there is a country stereotypical of regional inequality, it is Italy. The relatively late unification of the country and the complex history of the peninsula have created a well known narrative about the prosperous north and the pauper south. And, as a matter of fact it is true that average income in the North is substantially higher than in the South. The richest province having a GDP per capita around three times higher than the poorest. Consequently, one could think that the place of birth is an important determinant of the income of Italians and, consequently, that those born in provinces with higher average incomes play the lottery of life with better cards than those born in the southern provinces. The main result of our paper is to show that this is not the case. Geography plays only a marginal role in determining the inequality between Italians.

Yes, the North is richer (we will even show that the distribution of lifetime income in the North first order stochastically dominates that of the South), but the difference in average income between the rich and poor provinces is much smaller than the differences between individuals who live in any given province. That is: there are many very rich people in the South and many very poor people in the North. These differences within provinces are so large that in the general lottery of life, geography (the difference between provinces) is an irrelevant factor.

The absence of a geographical gradient to the structure of inequality in Italy is a fact not only in the cross-section but also, and more importantly, in the lifetime income of individuals. We calculate the lifetime income of a cohort of Italians (born in 1960) and show that knowing the province where a person was born (or resides) is essentially useless in trying to determine his or her income. Moreover, the reverse is also true. If you aim to guess the province where a certain Italian was born (or resides), knowing his or her lifetime income is almost useless.

This result may be surprising or not (we surely find it so), but it is by no means obvious. Education, gender, and sector of activity are also salient characteristics of individuals, as is the place of birth or residence. Our result is not obvious because these variables do help to predict income to a much better degree than geography does. Brunello et al. (2012) shows that industry explains approximately 25% of differences in wages, while education explains approximately 16% of differences in wages and 20% of differences in lifetime earnings. In contrast, we show that geography accounts for a mere 1.8% of the variance of lifetime income of men. Education, gender, and sector of activity are much more important drivers of inequality, yet the discussion of inequality in (and about) Italy is dramatically fixated on geographical differences.

Still, the fact that geography is an irrelevant driver of inequality in Italy has an important qualification. We will show that for women, but not for men, there are very large differences in participation that can be attributed to the province of birth. These differences in participation between provinces translate to differences in female average income between provinces that are able to explain a much larger share of the inequality between Italian women. In other words: the only way in which geography acts as a driver of inequality in Italy is that it helps predict female labor force participation and, thus, income. The effect on the overall population is small, but it explains about 10% of the differences in lifetime income between women.

The remainder of the paper is organized as follows. Section 2 summarises the literature on the topic. Section 3 describes the data. Section 4 presents evidence on cross-sectional inequality. In Section 5 we present lifetime inequality and our main results. In Section 6 we perform an experiment called "guess the province". Finally, Section 7 concludes and presents a general discussion.

2 Literature Review

The recent increase in within-country inequality is considered one of the greatest challenges of our times (Atkinson et al. 2011, Piketty and Saez 2007, Acemoglu and Autor 2011, Katz and Autor 1999, Piketty 2018). Higher levels of inequality threaten economic stability and can foster greater social and political instability (Galbraith 2012) raising concerns over the loss of upward mobility and declining opportunities for future generations in the United States (Sitaraman 2017), UK and elsewhere (Peck 2016).

The literature on earnings inequality and its long-term determinants is abundant, and within this literature many studies have focused on the analysis of the role of employers. Many articles have investigated the contribution to inequality of sorting, segregation, and pay premia (Song et al. 2019, Haltiwanger et al. 2022, Card et al. 2018, 2013). There is evidence that some firms pay workers with similar skills more than others (Krueger and Summers 1988, Van Reenen 1996) and, controlling for differences in observed and unobserved worker characteristics between firms, they have described how these differences in wage premia contribute to the distribution of earnings (Abowd et al. 1999, Goux and Maurin 1999, Abowd and Kramarz 1999, Holzer et al. 2011, Alvarez et al. 2018, Card et al. 2013). However, less attention has been paid to the regional dimension of earnings inequality. Among the few articles in the literature, Florida and Mellander (2016) examine the geographic variation across US metros distinguishing between wage and income inequality. They find that wage inequality is closely associated with skills, human capital, technology and metro size, while these factors are only weakly associated with income inequality. For the case of Canada, Breau (2015) shows that labor market, socio-demographic and institutional variables are key factors explaining differences in increasing regional inequality. Moser and Schnetzer (2017) find a strong positive correlation between regional income levels and inequality in Austria, where high-income municipalities exhibit a larger spread in the income distribution. For the case of Italy, the paper by Acciari et al. (2013), one of the first to investigate the spatial dimension of inequality in the country, uses tax records from 2000 to 2011 to compute the Gini coefficient for all Italian provinces. They show that inequality was higher in the South due to a smaller share of income held by the lower tail of the distribution, while higher in major metropolitan areas. Over time, inequality increased, particularly during the Great

Recession, due to a reduction in income, mainly among individuals with below the median income. These results are in line with the more recent findings of Acciari et al. (2021), who single out Italy as one of the counties with the strongest decline in the wealth share of the bottom 50% of the adult population. Using Italian social security data, Belloc et al. (2018) compute the within-between area variance decomposition of nominal and real wages in 2005. They find that around 95% of the variance is due to a within dimension, regardless of whether the within dimension refers to macro-regions, regions, or provinces. Then, they estimate no urban/rural wage premia for employees, subject to collective bargaining, while a large premia for self-employed individuals, not subject to collective bargaining. Boeri et al. (2021) show that Italy exhibits limited geographical wage differences in nominal terms, due to the nationwide sectoral contracts, which are binding and allow only for limited local wage adjustments. However, when taking into account inflation, wages turn out to be higher in the South, where productivity is lower, compared to the North, where both productivity and employment are higher. None of the papers in the literature however computes lifetime income, while they focused only on cross-sectional income.

3 Data

We use two main sources of data both provided by the Italian Social Security Institute (INPS), one of the largest administrative organizations at the European level. The first source is a longitudinal administrative employer-employee dataset that collects data on the working histories of the universe of private sector employees in Italy, who represent more than 70% of Italian workers. The data are structured as an unbalanced longitudinal sample at the individual (and firm) level at a yearly frequency. Together with earnings and employment histories, the INPS data include socio-demographic information regarding age, sex, nationality as well as province of birth, and of residence of both individuals and firms.

The second data source collects the records of all social security contributions ever paid by workers and by firms on behalf of the workers for a sample of individuals who represent approximately 13% of the whole population (Social Security Histories). In terms of labor outcomes, the dataset contains information on earnings and all types of benefits ever received by the individuals, including maternity and paternity leave benefits, unemployment allowance, sick leave benefits, short-term work programs (STW). We are able to infer whether the income the worker received came from her occupation as an employee in the private sector, in the public sector or whether she was self-employed when she paid the contributions.

While the first dataset has the advantage of collecting information on the universe of private sector employees, it does not include public sector, self employed, agricultural workers and caretakers and does not contain information about benefits, but only on earnings. The second dataset includes all records of social security contributions (wages and benefits), and all individuals independent of the job setting (private, public or self-employed), and it represents 13% of the population. We use the first data source to decompose inequality at cross-section level, while we use the second dataset to compute the lifetime income.

Specifically, in order to compute lifetime income, we focus on the cohort of individuals born in 1960: these workers were fifteen years old in 1975 (the year in which the data are available), and 56 years old in 2016 (the year in which our analysis terminates). This selection leaves us with a total of 113,388 individuals. We exploit the information on the province of birth in order to assign workers to geographical areas. In order to control for migration issues and for robustness purposes, we also use the last province where the contributions were paid to decompose inequality. It is important to mention that this dataset includes only individuals who have worked at least one day or have received some benefits in their lifetime, while individuals who have never worked do not show in the records.

Table 1: Decomposition of (log) income variance. Cross-sectional data.

	Total variance	Between variance	Within variance	Between share	Within share
1985	0.190	0.007	0.182	**3.93**	96.07
2018	0.267	0.008	0.259	**3.02**	96.98
Change	0.077	0.001	0.077	-	-
% Total increase	100.00	0.77	99.23	-	-

4 Cross-Sectional Inequality

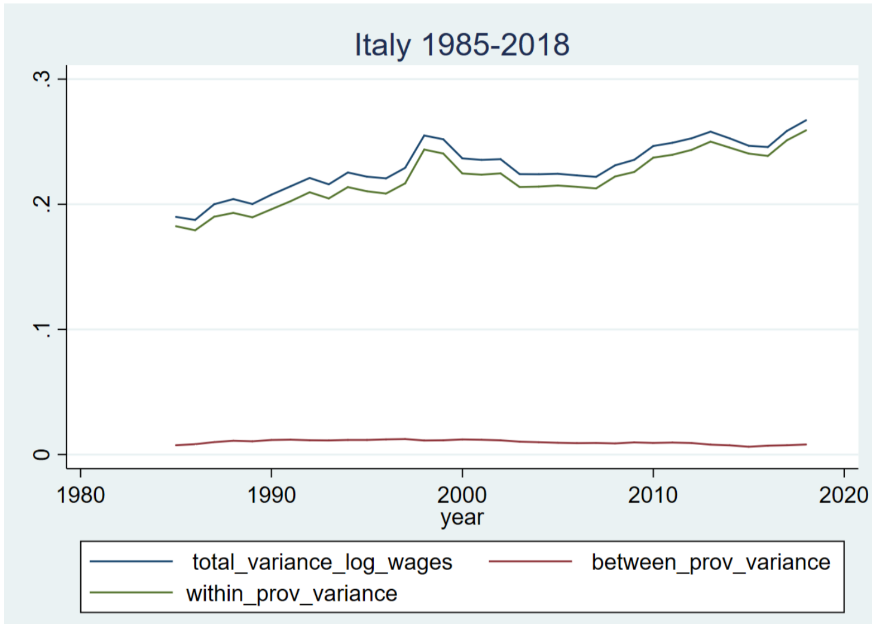
In order to assess how much geography matters in explaining inequality in Italy, we start by performing a simple variance decomposition. We use the INPS matched employer-employee dataset, from which we exploit information on wages and earnings. We select a cross-section of individuals in one year, we compute the variance of log wages in the given year and then we decompose the overall variance into two components: within- and between-province dispersion. Let w_{ip} be the log of the wage earned by individual i born in province p and let $\bar{w}$ be the average wage in Italy, we compute:

$$\underbrace{\frac{1}{N}\sum_{\forall i}(w_{ip}-\bar{w})^2}_{\text{total variance}} = \underbrace{\sum_{\forall p}\frac{n_p}{N}(\bar{w}_p-\bar{w})^2}_{\text{between province variance}} + \underbrace{\sum_{\forall p}\frac{n_p}{N}\frac{\sum_{\forall i|i\in p}(w_{ip}-\bar{w}_p)^2}{n_p}}_{\text{within province variance}},$$

where w_p is the average income in province p, N is the total population in Italy and n_p is the population in province p. This equation provides a simple way to decompose the total income dispersion in the economy into the between-province component (variability of average income across provinces) and into the within-province component (weighted average of within province dispersion using population shares as weights).

Table 1 reports the total variance and its decomposition in 1985 and 2018. In approximately 30 years, the total income variance increased by 40% from 0.19 to 0.267, however, the

Figure 1: Evolution of Cross-Sectional inequality of privately employed individuals and its decomposition in between and within provinces



between province component has played a negligible role, while a large increase is ascribable to the change in the within province component. Specifically, the between province share accounted for 3.93% of the total variance in 1985 and for 3.02% in 2018. The increase in the within-province variance accounted for 99.23% of the increase in the total variance in the period considered. This is represented in Figure 1, which shows a seemingly flat line representing the between province variance in the period 1985-2018, while increasing lines representing both the total and the within province variance.

Note that the share of the between province variance in total variance is equivalent to the R^2 of regressing individual income on provincial dummies. Thus, these are the results of performing regressions with 104 dummies (one per province), that yield a R^2 of about 3% in 2018. A way to shed light on the irrelevance of geography in explaining inequality is to compare it with the R^2 of alternative regressions. With our data, it is immediate to regress annual earnings to the sector of the firm. When we use NACE sector of activity at 2 digit level (88 categories), and nothing else, as an explanatory variable, we obtain a R^2 of 25% in

2018. This is, ordering people per sector of activity (even if using a 20% less of dummies) one is able to explain one order of magnitude more than when ordering people by province. Sector of activity gives order and structure to the data, in this sense explaining it; geography does not.

Thus, geography seems an irrelevant driver of cross-sectional inequality: average income might be different in the North and in the South, but the differences within each geographic area dwarf that difference in averages. Nevertheless, inequality in the distribution of cross-sectional income at a moment in time may not be the relevant variable to consider for our problem.

Firstly, because the dataset we use is by definition restricted to people who work, while unemployment rates are different in southern and northern provinces. It might be the case that the accumulated lifetime income is actually much lower in southern provinces, as their inhabitants suffer unemployment spells with more frequency. Thus, geography could be a bigger driver of inequality if we consider a lifetime notion of income.

The second reason is that while the variance of cross-sectional income needs to be larger than the variance of lifetime income (insofar as there is a mean-reverting component in the income processes that agents face) the mapping between both does not need to be the same between provinces. Imagine a province where the unemployed are always the same people and another where the people who suffer unemployment change over time. In the first province lifetime inequality would be larger than in the second, even if average unemployment is equal in both provinces.

In addition, it is self-evident that inequality in lifetime income is a better account of the inequalities in welfare than cross-sectional inequality. It is a better account of how differently life treats different individuals on top of and beyond the serendipity of small, passing, and ultimately irrelevant vicissitudes. Thus, we turn to the measurement of lifetime income in the next section.

5 Inequality of Lifetime Income

We want to abstract from issues of heterogeneity between cohorts. Thus, we focus on the cohort of individuals born in 1960, who were fifteen years old in 1975 (the first year in which the data are available), and 56 years old in 2016 (the year in which our analysis terminates). This leaves us with a total of 113,388 individuals.[1]

We compute the lifetime income, using the INPS dataset, which includes all the social security contributions of a representative sample (13%) of the Italian population. The lifetime income is computed as the logarithm of the sum of all income received during the individual working life. In each year we compute the real value of income using the annual national CPI. Following the literature (Song et al. 2019), we do not use an interest rate to compute the present discounted value of income. In years in which no income is recorded we manually add that year's equivalent of one euro in 2016, in order to obtain a balanced panel.[2]

We use four definitions of lifetime income. The first one sums only income coming from private sector employment over an individuals life. The second one sums income from both private and public sector employment. The third one sums all earnings, thus it also includes income from self-employment. Finally, in our most comprehensive definition, we also include all benefits received by the individual over their life, such as maternity and paternity benefits, unemployment subsidies, sick leave benefits etc.

Finally, we exploit the information on the province of birth in order to assign workers to geographical areas. In order to control for migration issues and for robustness purposes, we also use the last province where the contributions were paid to decompose inequality. We use the same decomposition as in Equation 1, substituting annual earnings for lifetime income.

[1]Remember that this database consists of a random sample of 13% of the population, and consequently, of the 1960 cohort.

[2]We do this because some individuals in the dataset have received no income in certain years (and we want to take logs). Moreover, as it will be clear later, we also generate artificial data for control with bias of representation in the data base across genders for those individuals with no participation.

Table 2: Decomposition of variance of (log) lifetime income.

	Total variance	Between variance	Within variance	Between share	Within share
All Sources. Province of Birth	11.040	0.377	10.663	**0.034**	0.966
All Sources. Province of Residence	11.029	0.468	10.560	**0.042**	0.958
No benefits. (inc. Self-employed)	12.271	0.547	11.724	**0.045**	0.955
Private and Public Employment	12.273	0.547	11.726	**0.045**	0.955
Private Employment only	14.144	0.752	13.393	**0.053**	0.947

Table 2 reports the total variance of the lifetime income of the cohort of 1960 and its decomposition between and within provinces. In different rows, we assign individuals to either their province of birth or residence, and we account for different income sources. The overall result is clear: geography is a marginal driver of inequality. Nevertheless, the different exercises teach some interesting lessons.

In the first row we assign individuals to their province of birth, and the between share component accounts for only 3.4% of the variance, but in the second row we assign them to their province of residence (the last province where they received income) and the between share component accounts only for 4.2% of the total variance. Thus, not only is the province of birth irrelevant in the overall picture of inequality, but it is also clear that this is not because of migration.

In the following rows, we always assign individuals to their province of birth but consider different sources of income in our measure of lifetime income. Including earnings from public sector employment naturally decreases total variance of lifetime income and including benefits decreases it even further, confirming the role of the welfare state in reducing inequality, while the inclusion of income from self-employment does not change things markedly[3]. In any case, in our focus of interest, when we decompose income into the two components, across all four income structures, we find that the between province share ranges between 3.4% and 5.3%: the vast majority of the income variance is to be found within provinces irrespective of income source. The share of the between province variance is slightly smaller when including transfers and income from public employment, but even when looking only at private sector earnings, the rather limited role of geography in accounting for total inequality of lifetime income is very clear.

5.1 Gender and Geography

Interestingly, the role of geography is more prominent when looking at differences across genders. In the first two rows of Table 3 we present the decomposition of lifetime income for men and women separately. Inequality of lifetime income is higher among men than among women, but the share that is explained by province is much higher among women. Specifically, while the between-province share is 7.1% among women, it is only 1.8% among men. The reason, of course, is that in the South women's participation is substantially lower than in the North, thus resulting in an observable driver of inequality: knowing the province where a woman was born helps predict her degree of participation and, thus, her income.

Actually, these numbers are an underestimation of the role of geography in accounting for the lifetime income of women. So far we have measured inequality by considering the individuals that appear in our dataset. This includes all individuals who have paid a social security contribution at least once in their lifetime (which could be a voluntary contribution to be eligible for pensions). However, we are fully aware that there are many inactive individuals who have never worked in their lifetime, and are not registered in the social security records. Interestingly, these people are not uniformly distributed, neither across

[3] Albeit the incomes reported by the self-employed may not reflect reality, as they are self-assessed and may be included with the only aim to be an investment in pension.

Table 3: Decomposition of (log) lifetime income variance by gender, assigning individuals to their province of birth. In the balanced sample, the number of females is artificially increased to be equal to the number of males.

	Total variance	Between variance	Within variance	Between share	Within share
Original data					
Males	11.169	0.198	10.971	**0.018**	0.982
Females	10.758	0.929	9.993	**0.071**	0.766
Balanced sample with "artificial women"					
Females	14.790	1.504	13.286	**0.102**	0.899

Italy nor across genders. The percentage of males in our data is larger in the southern provinces and correlates very negatively with the average lifetime income of the province (point correlation of -0.48).

To account for this phenomenon, we manually add "artificial" females to our data so that in each province we have a balanced sample of men and women, i.e., 50% men and 50% women. To keep coherence, we attribute to these "artificial" individuals an income equivalent to one 2016 Euro in each year of their life. We show the results in Table 3. Obviously, for men nothing changes, but in the third row of Table 3 we report that for women the role of geography increases: the between-province share rises to 10.2% of total variance.

Thus, we can summarize what we have learned so far:

1. In the context of total inequality in Italy, geography has only a marginal role. There are differences between the North and the South, but the differences within each province are vastly larger than any difference between provinces: knowing the province where a male was born does not help predict his income.

2. For women geography has a larger (albeit by no means predominant) role. This is because female participation is substantially lower in the South.

Our claim that geography is not an important driver of inequality has so far been based on the fact that we can not predict income by knowing the province individuals come from. In the next section we will perform the opposite experiment (try to guess the province knowing the income) to insure that our claim is correct.

6 Guess the Province

In this section, we perform the opposite experiment to what we have done so far. Instead of asking how much we know about the income of a person if we know his or her province of birth, we ask what is the probability of guessing the province of birth of a person correctly when knowing his or her income. It is another way of understanding the role of geography in accounting for income inequality.

Imagine a game called "Guess the Province". One province is drawn out of the 104 Italian provinces where each one of them has the same probability of being selected [4]. The game consists of guessing which province has been drawn. In the absence of any additional information, the probability of getting it right is exactly $\frac{1}{104}$, a bit less than 1%. Knowing the lifetime income of one person drawn randomly from the population of the province might in principle make the guess more accurate. The exercise consists of measuring how much better at guessing the province we get by learning about the lifetime income of people randomly drawn from the province. If that number is high, geography would be a very important driver of income. If it is low, it is an indication that it is not.

Notice that the posterior probability that this person is from province $\tilde{p}$ is:

$$P(p = \tilde{p}|w) = \frac{P(w|\tilde{p}) \times Q(\tilde{p})}{\sum_{\forall p} P(w|p) \times Q(p)}. \tag{1}$$

where $Q(p)$ is the prior that the province drawn is p (in our case $\frac{1}{104}$, but in principle this

[4] This is for simplicity, the probability could be proportional to population, or size, or arbitrary.

could be different[5]) and $P(w|p)$ is the distribution of income in province p.

People would guess the province with the maximum posterior probability.

$$\tilde{p} = \arg\max_{\forall p} P(p|w) \tag{2}$$

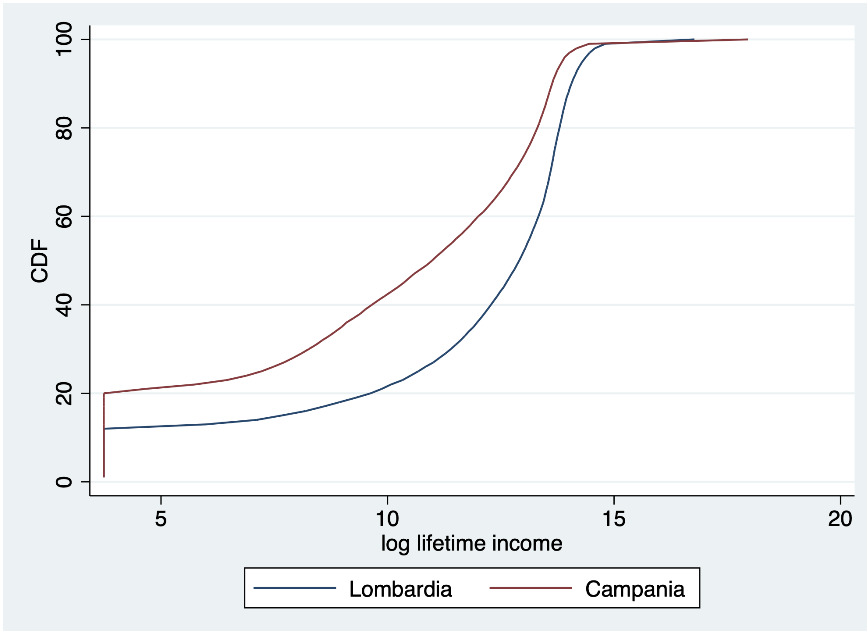
We simulate the game and calculate the percentage of times people get it right and compare it with the percentage of times people would get it right by randomly guessing. We define the success rate as the probability of guessing the province of birth correctly. Without knowing the lifetime income, the success rate is 0.97%, as there are 104 provinces with equal probability. Knowing one observation of the lifetime income, the success rate is 2.2%. That is, knowing one extraction of lifetime income it is possible to get the province right 2.32 times more often than in the scenario with no information, but still, that percentage is very low.

Not surprisingly, when we perform the same exercise by gender, we get something similar to our previous results. Having one observation of lifetime income and the additional information that the person is male, the success rate is 2.65%. The success rate in the case the person is female is 2.83%. That is, knowing one extraction of lifetime income of a male the probability to get the province right is 2.76 times higher than by guessing randomly, and 3.03 times if it is known that the person is a woman. It is 3 times better than when there is no information whatsoever, but yet 97.2% of the time the guess is wrong.

Thus, knowing the income of a person does not help in the "guess the province" game, confirming once again that geography is a marginal driver of differences in income among Italians.

[5] For instance, if provinces with larger population were drawn more often this prior should reflect that probability. We have performed these kinds of experiments and the results are always qualitatively identical.

Figure 2: CDF of the distribution of lifetime income in Lombardy and Campania.

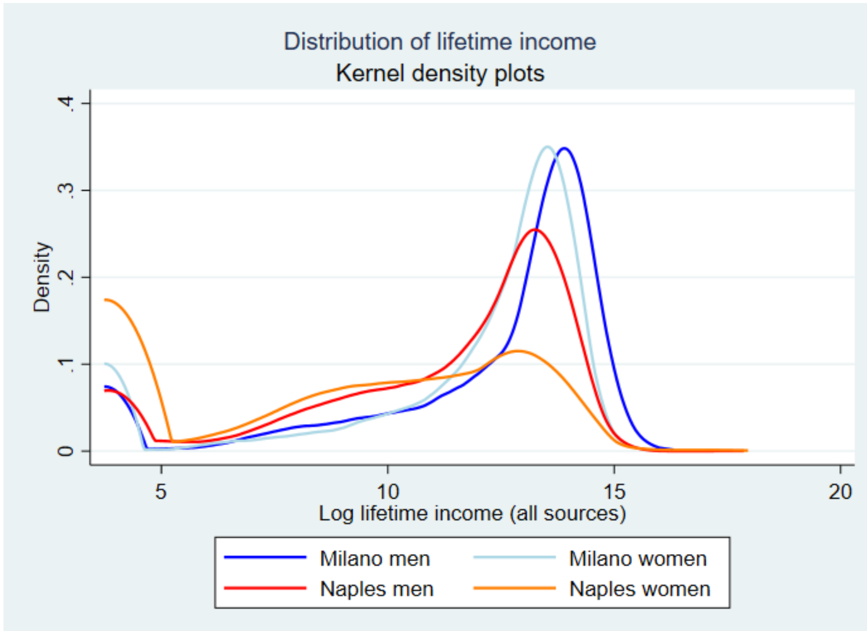


7 Conclusions and discussion

Using administrative data for Italy, we have shown that the vast majority of income inequality occurs within provinces, while the between-province component has only a marginal role. Nevertheless, this does not mean that geography has no role in income. It obviously does. The North is richer than the South. The interesting point is, we believe, that in terms of *dispersion*, the role that it plays is minimal when placed in the right context.

Figure 2 plots the CDF of lifetime income in two Italian regions, Campania and Lombardy, the stereotypical poor and rich regions in the country. The distribution of income in Lombardy essentially first order stochastically dominates the distribution of income in Campania. Without knowing more, and in the hypothetical scenario of being offered where to be born, it seems like a good idea to choose Lombardy. Our point is not that geography is irrelevant by itself, but that it is essentially irrelevant *for explaining the differences between Italians*. This is because, despite a clearly superior distribution of income in the North, the variance within each is so much larger than the difference in the averages, that in the lottery

Figure 3: PDF of distribution of lifetime income in Naples and Milan for males and females.



of life the issue of being born in one place or the other becomes almost irrelevant.

Perhaps the best way of visualizing this is to plot the Kernel density of lifetime income in both the North and the South. In Figure 3 we plot them for Milan and Naples (the capitals of Lombardy and Campania), separately for males and females. The spread of income in all four distributions is vastly larger than the differences in the averages. Although the averages are different, the average income of active women in Milan is higher than the average income of males in Naples. Still, the critical point that we are making is that there are many poor people in Milano and many rich in Napoli. The spread of any of the distributions is much larger than the difference between their means.

Consider now two lotteries. In the first one, the geography lottery, two tickets are available, "Naples" and "Milan", and conditional on the ticket you have, an income will be drawn from the lifetime distribution of the corresponding province.

The second lottery is the "relative income" lottery. There are also two tickets, they are called "poor" and "rich". Regardless of which of the two tickets you have, one of the provinces will be randomly drawn for you, and then if your ticket says "poor", you will get the income of a poor person in your province (say, the income of a person in the bottom 10%). If your ticket says "rich", the income of the top 10% in the province will be given to you.

Our point is that if you are playing the “relative income lottery”, you should be willing to pay a lot for the ticket "rich", but if you were playing the “geography” lottery, you should not be willing to pay much for the ticket "Milan".

As we have seen, there is a role for geography in the determination of female labor participation, but otherwise (and most certainly for men) in the big lottery of life the effect of being born in any province is marginal, almost irrelevant, at least when placed in comparison with the uncertainty of other aspects, such as being born in a relatively well-off family, having better education relative to others in the same province, or having better luck in finding the first job. Those uncertainties, the within-provinces serendipity, are to a much larger extent what determines an individual’s overall welfare. Not geography.

References

Abowd, J. and F. Kramarz (1999). *The analysis of labor markets using matched employer-employee data*. Handbook of Labor Economics, Volume 3, Volume 3B, Chapter 40. North Holland, Amsterdam.

Abowd, J., F. Kramarz, and D. Margolis (1999). High wages workers and high wages firms. *Econometrica 67*, 251–335.

Acciari, P., F. Alvaredo, and S. Morelli (2021). The concentration of personal wealth in italy 1995-2016. *Centre for Economic Policy Research, London.*

Acciari, P., S. Mocetti, et al. (2013). The geography of income inequality in italy. Technical report, Bank of Italy, Economic Research and International Relations Area.

Acemoglu, D. and Autor (2011). *Skills, tasks and technologies: Implications for employment and earnings.* Handbook of labor economics, vol. 4, ed. David Card and Orley Ashenfelter. Amsterdam: Elsevier.

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018). Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics 10*(1), 149–189.

Atkinson, A., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature 49*(1), 3–71.

Belloc, M., P. Naticchioni, and C. Vittori (2018). Premio salariale urbano, costo della vita locale e contrattazione collettiva. *VALUTARE CON I DATI AMMINISTRATIVI: PROGETTI VISITINPS SCHOLARS*, 10.

Boeri, T., A. Ichino, E. Moretti, and J. Posch (2021). Wage equalization and regional misallocation: evidence from italian and german provinces. *Journal of the European Economic Association 19*(6), 3249–3292.

Breau, S. (2015). Rising inequality in canada: A regional perspective. *Applied Geography 61*, 58–69.

Brunello, G., G. Weber, and C. Weiss (2012). Books are forever: Early life conditions, education and lifetime income. *Education and Lifetime Income*.

Card, D., A. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), 13–70.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics 128*, 967–1015.

Florida, R. and C. Mellander (2016). The geography of inequality: Difference and determinants of wage and income inequality across us metros. *Regional Studies 50*(1), 79–92.

Galbraith, J. K. (2012). *Inequality and instability: A study of the world economy just before the great crisis*. Oxford University Press.

Goux, D. and E. Maurin (1999). Persistence of interindustry wage differentials: A re-examination using matched worker-firm panel data. *Journal of Labor Economics 17*, 516–521.

Haltiwanger, J. C., H. R. Hyatt, and J. Spletzer (2022). Industries, mega firms, and increasing inequality. Technical report, National Bureau of Economic Research.

Holzer, H. J., J. I. Lane, D. B. Rosenblum, and F. Andersson (2011). *Where are all the good jobs going?: what national and local job quality and dynamics mean for US workers*. Russell Sage Foundation.

Katz and Autor (1999). *Changes in the wage structure and earnings inequality*. Handbook of labor economics, vol. 3, ed. Orley Ashenfelter and David Card. Amsterdam: Elsevier.

Krueger, A. and L. Summers (1988). Efficiency wages and the inter-industry wage structure. *Econometrica 56*(2), 259–293.

Moser, M. and M. Schnetzer (2017). The income–inequality nexus in a developed country: Small-scale regional evidence from austria. *Regional Studies 51*(3), 454–466.

Peck, J. (2016). Macroeconomic geographies. *Area Development and Policy 1*(3), 305–322.

Piketty, T. (2018). Capital in the twenty-first century. In *Capital in the twenty-first century*. Harvard University Press.

Piketty, T. and E. Saez (2007). Income and wage inequality in the united states, 1913-2002. *Top incomes over the twentieth century: A contrast between continental European and English-speaking countries 141*.

Sitaraman, G. (2017). *The crisis of the middle-class constitution: Why economic inequality threatens our republic*. Vintage.

Song, J., D. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming up inequality. *The Quarterly Journal of Economics 134*, 1–50.

Van Reenen, J. (1996). The creation and capture of rents: wages and innovation in a panel of uk companies. *The quarterly journal of economics 111*(1), 195–226.