Statistical Learning & Large Data project:

A Supervised Learning assessment of income mobility in Italy

Luna Boiago, Chiara Ferrara, Mariachiara Mariani, Anita Sammarini

17 May 2023

Contents

1 Introduction

Distribution and variation of income are important indicators of quality of life in a given country. While a static analysis of inequality is crucial to understand the drivers and factors contributing to income disparities in a society, a dynamic assessment of income mobility in a country is key to measuring the extent to which the socioeconomic spectrum of a given society is fair and fluid.

This work contributes to the literature on income inequality and mobility by providing estimates on the distribution of income in Italy from 2015 to 2019, employing data on the Italian Statistics Income Register provided by ISTAT, and examining the factors enhancing or hampering intragenerational income mobility.

The backbone of the analysis is structured according to a dual approach to income mobility. First, intragenerational mobility in Italy is assessed in a four-year time frame by means of Markov matrices conditional on respondents' demographic characteristics. Transition matrices are constructed using 2015 as initial year, and 2019 as final year, this allows to assess whether individuals shifted to a different income class four years after 2015. Mobility indexes derived from the transition matrices allow to evaluate and compare mobility across deciles and quartiles of the income distribution, along different societal classes.

In general terms, we are interested in investigating how individuals' income changes over time in Italy, the degree to which income mobility dynamics are heterogeneous across the population and which categories of society suffer most from a stagnant income mobility. While concentrating on mobility, this study builds on existing literature to further examine the North-South gap in Italy, as well as gender disparities, in a revised manner. The second part of the analysis builds on mobility indexes to expand on drivers of intragenerational mobility in Italy as well as factors that are, instead, hindering it. This study also develops on differentials in income mobility in Italy by closely examining significant factors contributing to the most stagnant or flexible income dynamics.

1.1 Mobility: definition and desirability

In this paper, we examine the first conceptual definition of **mobility**, positional change. It is thus important to outline some of its distinctive characteristics. This definition assumes that mobility for any specific individual depends on other people's positions as well. Each person's initial and final position depends on the positions of everyone else in society thus defining a hierarchy of positions. Therefore, if one person in the observed society changes position, the same must happen to at least another person. It is not possible for everyone to be upwardly mobile or downwardly mobile.

To fulfill the purposes of this study, it is important that we provide some background on the desirability of social mobility. To review the value of income mobility we must attempt to address the following questions: is income mobility socially desirable? Does a higher mobility signal a social improvement? Though this is neither the research question nor the main focus of this study, it is useful to consider the social meaning and context behind results.

Literature has provided an answer to these questions in a number of ways depending on the mobility concept employed and depending on whether within- or between- generation mobility was being assessed. Lower association between origin and destination positions have been associated to a more equal society. Lack of social mobility implies inequality of opportunity thus from this perspective greater mobility is socially desirable. (Aldridge, 2001)

This definition of an equal society is most appropriate when dealing with **intergenerational** mobility: an individual's opportunities in life should not depend on the socio-economic status of its parents. In an **intragenerational** context this definition loses strength. Nonetheless, even though income mobility has an inequality-reducing impact, mobility is not necessarily socially desirable if it represents transitory shocks since the transitory income is an idiosyncratic shock and as such greater variation implies greater risk and greater risk is undesirable for risk averse individuals.

Finally, when addressing the social desirability of individual income growth, there is no clear-cut answer. It could most generally be stated that an increase in income for any given individual is a social improvement and any fall is undesirable but when considering the welfare of society as whole, gains and losses of different individuals are subject to different weights. On the other hand, others might believe that differential income growth rates are not a concern if income gains among the rich correctly reward entrepreneurial activity.

1.2 Brief literature review

Recent literature has concentrated on country-specific studies on income inequality (Atkinson et al., 2011) however the Italian case deserves particular attention. Income inequality in Italy is comparatively high. Referring to the Gini coefficient as a measure of income inequality level, we can observe that Italy recorded a high-income inequality level during the late 1960s and the early 1970s followed then by a gradual improvement in the subsequent two decades. Since the 1990s, the Gini coefficient rose to 0.33 in 2000 and increased further to 0.35 in 2012 (Atkinson and Morelli, 2014). This rising trend is common in many of the OECD countries, nonetheless, unlike these countries, Italy's income inequality appears to be slightly above the OECD average.

The study of personal income distribution has its origin in the international debate ignited at the end of the 19th century by Pareto's analysis of the revenue curve. Since then, several studies in Italy have covered different aspects of income dispersion in the country and recent studies' main results from related literature are summarized as follows.

Between 1948 and 1968, the years of the Italian economic miracle, the level of income inequality did not undergo significant changes. However, these results are not subject to an easy interpretation since no comparable data is available for these years and it is thus not possible to state whether this trend is the result of relative stability. During the following decade, income distribution recorded increasing trends until the end of the 1970s. Since the 70s the Gini coefficient dynamics are W-shaped, displaying throughs in 1982 and 1991, and peaks in 1979, 1987, and 1995.

A strand of literature investigates inequality in the country by addressing the question from a

geographical angle. In general, these studies show that high disparities were found both among and within Italian macro-regions, with the highest top income concentrations in the North (Guzzardi et al., 2022). The literature has also investigated income inequality in a growth-related perspective. Kuznets (1955) and Barro (2000) have argued that the relationship between inequality and growth depends on the stage of economic development while the case of income inequality on economic growth in Italy has highlighted the significant negative impact of inequality on growth (Kirikos et al., 2017).

Moreover, recent literature investigating income disparities in Italy has found that gender income gaps follow a J-curved pattern, high at the bottom and even higher at the top of the income distribution (Guzzardi et al., 2022). Earlier research on income mobility has typically focused on either within or between-generation topics. The literature has focused on developing methods for the measurement of economic inequality (Jenkins and Van Kerm, 2006) and of intrageneration inequality and intertemporal mobility (Burkhauser et al., 2011). Mobility was found to be positively correlated with economic activity and social capital, and negatively correlated with inequality.

Within this context, the purpose of this study is to further develop the analysis on income dispersion in Italy. We refer to income inequality defined by means of individuals' disposable income, therefore income net of taxes plus subsidies. The main contributions of this analysis to the existing literature on Italian income distribution are: estimating the impact of education and gender differentials on an enhanced intragenerational mobility as well as describing differences in mobility dynamics across age classes and examining the impact of various sources of income on mobility (self-employment, employment, and pension income).

2 Methodology

In the present study, we have drawn data from the Italian Income Statistics Register (IISR) provided by the Italian National Institute of Statistics (ISTAT). This document consists in a collection of information about different forms of income, including in individual's gross and after-tax income. The sample includes 337,396 individuals and represents approximately 40 million taxable persons who reported their income in 2019. Each individual is associated with a coefficient representative of the relative weight of each individual in relation to the population: this means that sample weights indicate the number of individuals in the population represented by one individual in the sample.

The register covers a period that goes from 2015 to 2019 and provides longitudinal data on income records of the sampled individuals. Income records collected in the dataset represent income declared by individuals in their tax return (dichiarazione dei redditi), integrated with non-taxable-income traced by Istat. Non-taxable income refers to all those proceeds that are not subject to personal income tax (PIT), that is the Imposta sui redditi delle persone fisiche (IRPEF). Income revenues that are not subject to personal income taxation include payroll subsidies, unemployment benefits, tax-exempt pensions, reddito di cittadinanza, income for casual labour, fringe benefits and

subsidies to families. Reddito di cittadinanza is a social welfare system introduced in Italy which aims to guarantee a monthly allowance to integrate the incomes of families with limited economic resources.

Total gross income can be split into various components according to the type and source of income received. Pursuant to Art. 6 of the Income Tax Consolidation Act individual incomes are categorized as follow:

- Property Income
- Capital Income
- Employment Income
- Self-employment Income
- Business Income

We have chosen to work with the employment, the self-employment and the pension income. Employment income consists of gross earnings and employers' social contributions, whereas self-employment income is net income earned from your own trade or business.

Furthermore, the dataset reports a set of individual characteristics, gender, age class, education, Italian territory geographical partition in terms of macro-region, and region of residence. We have created dummy variables for the geographic partition, in so doing obtaining four partitions that provide a broad prospective on respondents' residence: North-West, North-Est, Center and South. Age is split into eight variables grouped by ten, except for the youngest and oldest individuals whose grouping is not symmetrical as the others. We have created a dummy variable for each age class. Finally, for the variable study title, there are two dummy variables: one for the individual with the lowest study title; the other one for the individual with the highest qualification. In this way, we pursue our aim of studying the income differences among individuals depending on their gender, their education level, their age, and their Italian geographical area of provenance.

The distribution of demographic variables is presented for the entire sample and for each quartile of the 2019 disposable income distribution. The first quartile refers to the lowest 25% of the distribution, thus the poorest quartile of the sample, and the fourth quartile contains information on the richest 25% of our sample. Through this subdivision of the distribution into percentiles and deciles, it has been possible to observe whether each individual was in the highest or in the lowest decile and percentile and if they switched from one to another from 2015 to 2019. Finally, we have also calculated the income shares of each individual in order to see how it was distributed among the employee compensation, the self-employment and pensions.

The methodology that has been used in this study is four parts. The first one is based on a multinomial logistic regression, in which we have used as independent variables the gender, the study title, the class age and the geographic partition, whereas the dependent variable is the quartile to whom is related an individual with his/her own specific characteristics in 2019. Through this

procedure, we wanted to collect data about the probability that individuals with particular features are in the four different quartiles. We have then created a dummy variable that is equal to 0 if the individual has switched to the richest quartile in 2019 and to 0 if the individual was already in the highest quartile in 2019 or if he/she passed to a lower one. We have repeated the same operation with the deciles. The second step of our methodology consists of four logistic regressions. In each of them, we have used as independent variables gender, class age, study title, geographic partitions, employee, autonomous and pensions income in 2019. For the first one, the aim was to estimate the probability of passing to the richest decile and has as dependent variable a dummy variable assuming value 1 if the individual belongs to the richest decile in 2019 and belonged to a lower decile in 2015; the second logit model has a dependent variable a dummy variable assuming value 1 if the individual belongs to the richest quartile in 2019 and belonged to a lower quartile in 2015 and it aims at estimating the probability of passing to the richest quartile. In the other two logistic regressions we have considered the poorest decile and percentile and in order to estimate the probability to pass to the poorest quartile and then to the poorest decile. The third pace was repeating the logistic regressions through delta regressions, using then discrete variables, which assume value from -100 to 100, with the goal of estimate how many steps – in terms of delta percentiles and delta deciles – each individual has done from 2015 to 2019. Finally, we have used a transition matrix both of deciles and quartiles in 2015 and 2019 for describing the probabilities of moving from one state to another in a dynamic system.

3 Regressions

3.1 Multinomial logistic regression

Using the above-mentioned multinomial regression we estimated the probability for individuals to be classified in each one of the four different quartiles knowing their age class, their gender, their study title, and the geographical partition to which they belong. For example, these are the predictions for a woman with a degree who lives in the South considering that her starting point in 2015 was the first quartile and considering all the eight age classes.

The probability that a woman remain in the first quartile in 2019 is always higher for each age class. This becomes a little lower, thus suggesting a slightly larger probability for individuals to be placed in a higher range, for women in their twenties and thirties. Considering the same age class again, the probability that a woman transits into a different quartile increases sharply if the starting one is the second. The value is in fact around 30% instead of the previous 60%. Starting from the third or fourth quartile, however, it becomes more difficult for a woman with those specific characteristics to change range. The probabilities that women stay in the third and fourth quartiles are about 60% and 70% respectively (always looking at twenty or thirty-year-olds).
Trying to change only the sex, it was noticed that men seem to have greater chances of transitioning

	First quartile	Second quartile	Third quartile	Fourth quartile
1	0.9543475	0.03425292	0.01139231	7.303057e-06
2	0.7652243	0.16330261	0.06486368	6.609412e-03
3	0.6100537	0.25442092	0.11612311	1.949227e-02
4	0.6657680	0.20204248	0.10193832	3.025136e-02
5	0.7170126	0.16457713	0.07643402	4.197623e-02
6	0.7777591	0.13521173	0.05208421	3.494494e-02
7	0.8122599	0.11141099	0.03809508	3.823401e-02
8	0.7863341	0.14998911	0.03507338	2.860339e-02

Table 1: Transition probability prediction of a woman with a degree in the south, whose starting quartile in 2015 was the first one

to a quartile different from that of 2015 except in the case that the latter was the fourth one.
Then we repeated the analysis taking into account people with a low degree. Wanting to make a comparison with the results reported previously, we can note that the probability that women in their twenties or thirties remain, for example, in the second quartile is lower but this does not lead to an increase in the possibility to improve their status because, among the other three, the one to increase is the probability of moving to the lower quartile which in this case is the first.
A further analysis was conducted looking at the difference that could result in being in the north instead of the south. If in the previous table the probability that a graduated woman from the south belonging to the third age class remained in the first quartile was equal to 61%, the same type of probability associated with a woman with the same characteristics but living in the north becomes 46%. The latter seems more likely to improve her condition just because of her geographical region.

3.2 Logistic regressions

The estimated regression models are the following:

$$\begin{aligned} Quantile_j = \beta_0 + \beta_1 Gender_j + \sum_{j=2}^{9} \beta_j AgeClass_j + \beta_{10} Degree_j + \beta_{11} NoStudy_j + \\ \beta_{12} EMP_j + \beta_{13} SEMP_j + \beta_{14} PEN_j + \sum_{j=1}^{14} \gamma_j Region_j + \\ + \sum_{j=1}^{4} \theta_j Quartile_j + \sum_{j=1}^{10} \xi_j Decile_j, \end{aligned} \tag{1}$$

where $Quantile_j$ can either be Quartile **1**, Quartile **4**, Decile **1** or Decile **10**, and is a binary variable assuming value 1 if the individual belongs to the chosen quantiles in 2019 given that he/she belonged to a different income class in 2015. With this dependent variable therefore I aim at capturing the probability to transition in time to one of the chosen four quantiles.
$Gender_j$ assumes value 1 if the individual is female. The variable $AgeClass_j$ is divided into eight different age classes and the its reference category is the age class ranging from 55 to 64. Predictors

$Degree_j$ and $NoStudy_j$ are binary indicators of educational attainment equal to 1 in the first case if the individual has earned a degree and equal to 1 in the second case if the individual does not possess any study title at all. Variables EMP_j , $SEMP_j$ and PEN_j are variables for individuals' sources of income representing the share of total income deriving from employment, income from selfemployment and pension income. $\sum_{j=1}^{14} Regione_j$ j=1 $Regione_j$ represents fourteen dummy variables for every italian region, in some cases regions are collected into couples. Marche-Umbria are considered together as well as Puglia-Basilicata, Calabria-Sicilia and Piemonte-Valle d'Aosta. The latter is chosen as baseline category for the interpretation and is thus excluded from the regression. Finally, in order to control for respondents' initial position in terms of the income class that individuals belonged to in 2015, binary variables for quantile income classes were inserted in the model.

The following table presents the estimates of equation (1) for each of the four regression models. First consider the model that aims at estimating the probability of passing to the richest decile. Women are less likely than men to move into the tenth decile in 2019 ($\beta_1 = -0.006$). The results change when considering the lowest quartile and decile. In the first case being a woman increases the chances of passing there while in the second the value of the coefficient is not statistically significant.

As concerning the age classes, there is a higher probability to pass to the richest decile if the age is between 15 and 54 with respect to the age class 55-64. The results are the same considering the probability of transiting to the lowest and highest quartile but change referring to the lowest decile. These age classes display a lower probability of moving to the poorest decile and a higher probability of moving both to the richest and to the lowest quartile. Coefficients from all four regressions are significant. Older individuals , aged 65 to 74 and older than 75, all display lower probabilities compared to the reference category of moving to the lowest quartile and decile. Values referring to the highest quartile and decile are not statistically significant

Having a degree increases the transition probability of moving to the higher bands and decreases the probability of reaching the poorer quartile. It seems to be easier to reach the poorest decile if you do not have a study title ($\beta_{11} = -0.026$)

Having a share of self-employment than not having is estimated to have a positive effect on the probability of transitioning both to highest and lowest quantiles, while a higher employment income share is estimated to have the exact opposite effect, decreasing transition probabilities.
The coefficients for the share of pension income suggest a decrease of the probability of passing to the highest decile and an increase when transiting to the lowest quantiles.

Among those that reflect a greater probability than Piemonte and Valle D'Aosta (reference regions) to transit towards the highest quantities there are Trentino Alto Adige and Friuli Venezia

Giulia. Coefficients of Lombardia, Emilia Romagna and Toscana are significant only referring to the 10^{th} decile. However, Calabria Sicilia reflect a lower probability than Piemonte and Valle D'Aosta to transit towards both the highest quartiles and deciles while a grater one considering the lowest quantiles. Campania, Puglia-Basilicata, Abruzzo-Molise and Lazio display higher probability to transit to lowest quantiles too These results are in line with expectations suggesting that for Southern regions a future worsening of economic conditions is more probable compared to Northern regions such as Piemonte and Valle d'Aosta.

Models (1) and (4) are particularly interesting when examining coefficients ξ_j that measure how starting from a given decile of the income distribution impacts on the probability of shifting in the future to the highest or lowest decile. In model (1) they are all significant and negative while they are all significant and positive in model (4).
This result suggests that belonging to any of the first eight deciles of the distribution lowers the probability of transitioning to the highest decile. This could indirectly be signaling the relative difficulty of reaching the highest income class since all starting positions, even the highest deciles, thus even the richest individuals, have a lower probability of reaching decile 10. On the contrary, coefficients for model (4) are all positive and high reflecting, instead, the relative ease of moving down to the lowest income decile given any starting point. Coefficients become gradually lower considering deciles in ascending order, individuals starting off in the poorest deciles display higher probability estimates than individuals starting off from richer deciles. An important remark to make is that coefficients ξ_j are assumed to be the same regardless of the indivual's gender, age class or region of origin.

Table 2: Regression analysis output of the four logit models

	Dependent variable:			
	10^{th} Decile	4^{th} Quartile	1^{st} Quartile	1^{st} Decile
	(1)	(2)	(3)	(4)
Gender	-0.006^{***}	-0.022^{***}	0.019^{***}	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Age 0-14	0.013^{*}	0.011	0.024^{**}	-0.183^{***}
	(0.007)	(0.009)	(0.010)	(0.009)
Age 15-24	0.028^{***}	0.043^{***}	0.044^{***}	-0.197^{***}
	(0.006)	(0.008)	(0.009)	(0.008)
Age 25-34	0.034^{***}	0.058^{***}	0.037^{***}	-0.189^{***}
	(0.006)	(0.008)	(0.009)	(0.008)
Age 35-44	0.027^{***}	0.043^{***}	0.035^{***}	-0.189^{***}
	(0.006)	(0.008)	(0.009)	(0.008)
Age 45-54	0.021^{***}	0.035^{***}	0.029^{***}	-0.192^{***}
	(0.006)	(0.008)	(0.009)	(0.008)
Age 65-74	0.005	0.009	-0.030^{***}	-0.246^{***}
	(0.006)	(0.008)	(0.009)	(0.008)
Age >75	0.0003	0.003	-0.051^{***}	-0.265^{***}
	(0.006)	(0.008)	(0.009)	(0.008)
Study title: Degree	0.024^{***}	0.055^{***}	-0.017^{***}	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
No study title	0.003^{*}	-0.001	0.006^{**}	-0.026^{***}
	(0.002)	(0.002)	(0.003)	(0.002)
Self-employment share	0.026^{***}	0.027^{***}	0.044^{***}	0.026^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Employment share	-0.026^{***}	-0.027^{***}	-0.044^{***}	-0.026^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Pension share	-0.011^{***}	0.008^{*}	0.020^{***}	0.013^{***}
	(0.003)	(0.005)	(0.005)	(0.004)

	Dependent variable:			
	10^{th} Decile	4^{th} Quartile	1^{st} Quartile	1^{st} Decile
	(1)	(2)	(3)	(4)
Region Lombardia	0.005^{***}	0.003^{*}	-0.003	-0.003^{*}
	(0.001)	(0.002)	(0.002)	(0.002)
Region Trentino Alto Adige	0.014^{***}	0.020^{***}	-0.004^{*}	-0.008^{***}
	(0.002)	(0.002)	(0.003)	(0.002)
Region Veneto	0.003^{*}	-0.003	-0.003	-0.004^{*}
	(0.002)	(0.002)	(0.003)	(0.002)
Region Liguria	0.002	-0.003	-0.008^{***}	-0.006^{**}
	(0.002)	(0.003)	(0.003)	(0.002)
Region Emilia Romagna	0.008^{***}	0.004	0.004	0.002
	(0.002)	(0.003)	(0.003)	(0.003)
Region Friuli Venezia Giulia	0.004^{**}	0.005^{**}	-0.001	-0.004^{*}
	(0.001)	(0.002)	(0.002)	(0.002)
Region Toscana	0.004^{***}	-0.001	0.0003	0.0002
	(0.002)	(0.002)	(0.002)	(0.002)
Region Marche-Umbria	-0.001	-0.008^{***}	0.004	-0.001
	(0.002)	(0.002)	(0.003)	(0.002)
Region Lazio	0.002	0.0004	0.017^{***}	0.006^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Region Abruzzo-Molise	-0.002	-0.011^{***}	0.024^{***}	0.007^{***}
	(0.002)	(0.003)	(0.003)	(0.003)
Region Campania	-0.001	-0.009^{***}	0.025^{***}	0.014^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Region Puglia-Basilicata	-0.003^{*}	-0.013^{***}	0.022^{***}	0.010^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Region Calabria-Sicilia	-0.003^{**}	-0.013^{***}	0.029^{***}	0.010^{***}
	(0.001)	(0.002)	(0.002)	(0.002)
Region Sardegna	-0.002	-0.017^{***}	0.013^{***}	0.001
	(0.002)	(0.003)	(0.003)	(0.003)

	Dependent variable:			
	10^{th} Decile	4^{th} Quartile	1^{st} Quartile	1^{st} Decile
	(1)	(2)	(3)	(4)
First decile 2015	-0.049^{***}			
	(0.001)			
Second decile 2015	-0.047^{***}			0.221^{***}
	(0.001)			(0.002)
Third decile 2015	-0.047^{***}			0.083^{***}
	(0.001)			(0.002)
Fourth decile 2015	-0.045^{***}			0.061^{***}
	(0.001)			(0.002)
Fifth decile 2015	-0.045^{***}			0.041^{***}
	(0.001)			(0.002)
Sixth decile 2015	-0.040^{***}			0.024^{***}
	(0.001)			(0.002)
Seventh decile 2015	-0.035^{***}			0.012^{***}
	(0.001)			(0.002)
Eighth decile 2015	-0.021^{***}			0.011^{***}
	(0.001)			(0.002)
First quartile 2015		0.040^{***}		
		(0.001)		
Second quartile 2015		0.050^{***}	0.219^{***}	
		(0.001)	(0.001)	
Third quartile 2015		0.140^{***}	0.057^{***}	
		(0.001)	(0.002)	
Fourth quartile			0.041^{***}	
			(0.002)	
Constant	0.046^{***}	-0.024^{***}	-0.017^{*}	0.225^{***}
	(0.006)	(0.008)	(0.009)	(0.008)
Observations	214,102	214,102	214,102	214,102
Log Likelihood	104,512.400	39,456.130	11,534.320	49,192.570
Akaike Inf. Crit.	$-208,952.900$	$-78,850.270$	$-23,006.640$	$-98,315.150$
Note:				*p<0.1; **p<0.05; ***p<0.01

3.3 'Delta' Regressions

The estimated regression models are the following:

$$\begin{aligned} DeltaQuantile_j = \beta_0 + \beta_1 Gender_j + \sum_{j=2}^{9} \beta_j AgeClass_j + \beta_{10} Degree_j + \\ \beta_{11} NoStudy_j + \beta_{12} EMP_j + \beta_{13} SEMP_j + \beta_{14} PEN_j + \sum_{j=1}^{14} \gamma_j Region_j + \\ + \sum_{j=1}^{4} \theta_j Quartile_j + \sum_{j=1}^{10} \xi_j Decile_j, \end{aligned} \tag{2}$$

$$\begin{aligned} RelativeDeltaQuantile_j = \beta_0 + \beta_1 Gender_j + \sum_{j=2}^{9} \beta_j AgeClass_j + \beta_{10} Degree_j + \\ \beta_{11} NoStudy_j + \beta_{12} EMP_j + \beta_{13} SEMP_j + \beta_{14} PEN_j + \sum_{j=1}^{14} \gamma_j Region_j + \\ + \sum_{j=1}^{4} \theta_j Quartile_j + \sum_{j=1}^{10} \xi_j Decile_j, \end{aligned} \tag{3}$$

where $DeltaQuantile_j$ is a continuous variable measuring the number of quantiles that a given individual in the sample was able to jump in the four year time frame (from 2015 to 2019). The dependent variable for these regressions was calculated as the difference between the percentile to which an individual belonged in 2019 and the percentile to which that same individual belonged in 2015. Thus, this variable is a measure of "how many percentiles richer or poorer" a given individual is, compared to its starting position in 2015. $RelativeDeltaQuantile_j$ instead is a normalized measure of the the previously mentioned delta variable as it is computed by dividing the delta by the individual's initial position:

$$\frac{(percentile_j^{2019} - percentile_j^{2015})}{percentile_j^{2015}}$$

First consider the delta regressions to estimate how many steps - in terms of delta percentiles and deciles - each individual has done from 2015 to 2019. Estimates suggest that being a woman is a associated with a falling in the difference between the 2019 and 2015 percentiles equal to -3.007 (-0.315 referring to deciles) with respect to men.
Individuals belonging to all age classes (excluding the reference one) are estimated to increase their position moving towards both richer percentiles and deciles than those aged between 55 and 64

years.
Those with a degree in 2019 rose by about 6 percentiles and 0.6 deciles compared to those without one. The difference between deciles and percentiles is also positive for those who do not have a study title.
Self-employment share leads to an increase of -1.574 percentiles and -0.184 deciles while, on the contrary, the impact of employment share is positive. Once again being in a southern region has a negative impact compared to being in Piemonte and Valle d'Aosta. Those living in Calabria-Sicilia, for example, in 2019 backwarded 2.545 percentiles and 0.270 deciles with respect to the reference regions. However, those living in Trentino Alto Adige forwarded 1.532 percentiles and 0.041 deciles.

Table 3

	Dependent variable:	
	Δ percentile	Δ decile
	(1)	(2)
Gender	-3.007^{***}	-0.315^{***}
	(0.070)	(0.007)
Age 0-14	12.944^{***}	1.325^{***}
	(0.688)	(0.070)
Age 15-24	16.076^{***}	1.724^{***}
	(0.628)	(0.064)
Age 25-34	13.411^{***}	1.487^{***}
	(0.623)	(0.063)
Age 35-44	11.990^{***}	1.353^{***}
	(0.621)	(0.063)
Age 45-54	10.936^{***}	1.260^{***}
	(0.621)	(0.063)
Age 65-74	11.570^{***}	1.329^{***}
	(0.620)	(0.063)
Age >75	12.568^{***}	1.424^{***}
	(0.615)	(0.062)
Study title	5.778^{***}	0.588^{***}
	(0.099)	(0.010)
No study title	1.050^{***}	0.081^{***}
	(0.174)	(0.018)
Self-employment share	-1.574^{***}	-0.184^{***}
	(0.099)	(0.010)
Employment share	1.529^{***}	0.181^{***}
	(0.098)	(0.010)
Pension share	-2.997^{***}	-0.270^{***}
	(0.345)	(0.035)

	Dependent variable:	
	Δ percentile	Δ decile
	(1)	(2)
Region Lombardia	0.378^{***}	0.041^{***}
	(0.135)	(0.014)
Region Trentino Alto Adige	1.532^{***}	0.154^{***}
	(0.170)	(0.017)
Region Veneto	0.129	0.016
	(0.175)	(0.018)
Region Liguria	0.417^{**}	0.042^{**}
	(0.196)	(0.020)
Region Emilia Romagna	0.129	0.011
	(0.207)	(0.021)
Region Friuli Venezia Giulia	0.522^{***}	0.058^{***}
	(0.153)	(0.016)
Region Toscana	0.275^{*}	0.026
	(0.164)	(0.017)
Region Marche-Umbria	-0.332^{*}	-0.036^{**}
	(0.170)	(0.017)
Region Lazio	-0.651^{***}	-0.074^{***}
	(0.161)	(0.016)
Region Abruzzo-Molise	-2.050^{***}	-0.208^{***}
	(0.201)	(0.020)
Region Campania	-2.124^{***}	-0.226^{***}
	(0.161)	(0.016)
Region Puglia-Basilicata	-2.154^{***}	-0.224^{***}
	(0.155)	(0.016)
Region Calabria-Sicilia	-2.545^{***}	-0.270^{***}
	(0.143)	(0.014)
Region Sardegna	-1.689^{***}	-0.181^{***}
	(0.215)	(0.022)

	Dependent variable:	
	Δ percentile	Δ decile
	(1)	(2)
First decile 2015	19.356***	2.073***
	(0.145)	(0.015)
Second decile 2015	13.006***	1.389***
	(0.139)	(0.014)
Third decile 2015	9.014***	0.961***
	(0.138)	(0.014)
Fourth decile 2015	6.639***	0.729***
	(0.135)	(0.014)
Fifth decile 2015	4.590***	0.514***
	(0.133)	(0.014)
Sixth decile 2015	3.057***	0.355***
	(0.132)	(0.013)
Seventh decile 2015	2.276***	0.275***
	(0.130)	(0.013)
Eighth decile 2015	1.197***	0.157***
	(0.129)	(0.013)
Constant	−17.615***	−1.972***
	(0.626)	(0.064)
Observations	214,102	214,102
R^2	0.139	0.144
Adjusted R^2	0.139	0.144
Residual Std. Error (df = 214066)	15.273	1.550
F Statistic (df = 35; 214066)	988.578***	1,029.205***
Note:	*p<0.1; **p<0.05; ***p<0.01	

3.4 Transition Probability Matrix

Income mobility can be summarized with a transition matrix.
Cell entries ajk are the probability that an individual income in quantile j in period 1 is found in quantile k in period 2. Maximum immobility occurs when every person has the same position in x and in y, when all individuals are on the leading diagonal of the transition matrix, $ajk = 1$ in all quantiles.
Transition probabilities displayed in the General matrix were obtained without conditioning on any additional factor and are thus the least accurate out of all the presented transition matrices. This means that cell entries of this transition matrix display spurious transition probabilities since they do not control for the variability explained by additional factors such as age, gender, and geographical region.
Simply observing proportions displayed in the matrix it is possible to notice that in every row, the highest percentages are on the diagonal, signaling that, in general, **most individuals did not change their condition over time**. The highest persistence is recorded for the highest income classes in both matrices. In fact, 77% of the individuals who belonged to the richest decile still do in 2019 and 82% of those who belonged to the richest quartile in 2015, are still positioned in the highest income class in 2019. There appears to be a significant number of individuals who shifted to the nearest worse quantile, approximately one quarter of individuals in every row has moved one position on the left. The number of individuals who instead moved one position to the right is much lower on average, the proportion of individuals who exhibit this kind of improvement ranges from 9% to 15%. Results are obviously dependent from the number of quantiles used. With deciles, a more fine-grained look at income mobility is possible whereas quartiles allow for a more general analysis. This kind of pattern, in fact, is not as visible when considering income distribution across quartiles, even though shifts to the left across rows are always higher in absolute value compared to shifts to the right. In the top far right corner we observe the proportion of individuals who were able to perform a significant improvement in their income class and in the bottom far left corner those who instead experienced a massive negative downturn in income.
These areas of the transition matrix display the lowest proportions, signaling minor jumps in transitions between classes. Throughout the rest of this section we will first look at these first two matrices which we will refer to as General, since these were computed without controlling for others factor that might be affecting mobility.

Table 4: Deciles and Quartiles transition matrix

					Last year					
First year	1	2	3	4	5	6	7	8	9	10
2015					**2019**					
1	0.56	0.13	0.09	0.07	0.05	0.04	0.02	0.01	0.01	0.02
2	0.22	0.47	0.11	0.08	0.05	0.03	0.02	0.01	0.01	0.01
3	0.07	0.27	0.40	0.09	0.08	0.04	0.02	0.01	0.01	0.01
4	0.05	0.06	0.28	0.37	0.09	0.08	0.04	0.02	0.01	0.01
5	0.03	0.03	0.05	0.28	0.37	0.10	0.07	0.03	0.02	0.01
6	0.02	0.02	0.03	0.06	0.27	0.37	0.13	0.06	0.03	0.01
7	0.01	0.01	0.02	0.03	0.05	0.26	0.40	0.14	0.06	0.02
8	0.01	0.01	0.01	0.01	0.02	0.05	0.25	0.46	0.15	0.03
9	0.01	0.00	0.01	0.01	0.01	0.02	0.04	0.23	0.56	0.12
10	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.15	0.77

4 LASSO

4.1 Theoretical framework

In order to avoid overfitting of the data and select our features, we implemented Lasso regression. The word "LASSO" stands for Least Absolute Shrinkage and Selection Operator. It is a regularization technique that allows to improve over Least squares reducing sample variability of the coefficient estimators. Regularization is implemented by adding a so called *penalty* term. The latter is equal to the absolute value of the magnitude of the coefficient.
The tuning parameter λ controls the level of penalty applied to the model. By increasing this value, it is possible to obtain sparse models with only a subset of variables involved. In fact, lasso regression shrinks the coefficient estimates towards zero, thereby removing useless variables from the equation.

The lasso coefficients minimize the following quantity:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{i,j})^2 + \lambda\sum_{j=1}^{p}|\beta_j| = RSS + \lambda\sum_{j=1}^{p}|\beta_j| \tag{4}$$

λ can be any value from 0 to positive infinitive and it is determined using Cross Validation
It is important to remember that as λ grows, the variance decreases but the bias increases (or it is created if LS is unbiased). The selection of an appropriate value is made minimizing the out-of-sample MSE. Different values of λ are evaluated to arrive at a final model that strikes a balance between model parsimony and accuracy.

4.2 Procedure followed

The LASSO can be performed on a chosen response variable to select the best predictors for that variable. Thus, we chose 4 different response variables and performed 4 different LASSO procedures in order to identify the set of predictors that best explains and fits all the 4 responses together. The response variables chosen are dummies assuming value 1 if they respect the following characteristics, 0 otherwise:

- the richest decile in 2019;
- the richest quartile in 2019;
- the poorest decile in 2019;
- the poorest quartile in 2019.

For each of these variables, we first performed a simple LASSO and plot it, to observe how the coefficients of all the predictors shrinked to 0. However, this result is not enough. Thus, we chose a cross-validation method to better assess the previous results and observe which value of

the parameter λ could work in order to choose the best set of predictors. With the R function *cv.glmnet*, we obtained two outputs:

1. the values of λ, corresponding to the optimal one and the optimal + one standard deviation;
2. the corresponding plot: on the X axis the log values of λ, while in the Y axis the Mean squared errors. Until the curve stays flat (in the range between the optimal and optimal + 1se of λ) then we can choose the corresponding number and selection of coefficients.

Given these cross-validation results, the next step consisted in identifying the matrix of regressors whose coefficients did not shrink to zero according to the lasso, for both the optimal and the 1 se values of λ. We also plotted the trend of the coefficients with the label of the associated variables and the vertical lines corresponding to the two threshold values of λ.

Using the models build as previously described, we tried to test the ability of such models in predicting the response variables. We obtained quite satisfying results in terms of variance explained (ranging from 20 to 60 % according to the different response variable).

The last part of this section implied finding the common set of predictors that were selected in all the four cases of the four response variables, to re-perform logistic regressions on these as dependent variables.

To do so, we first identified the selected predictors for each variable, using as λ the 1se. This choice is due to the fact that there is no significant variation in terms of MSE between the optimal value of the parameter and the 1se. Thus, it is more desirable to use the value such that a smaller number of predictors is selected.

A further observation is necessary: as can be seen in Figure 2, the curve of the MSE stays pretty flat also for values of λ greater than the one chosen. However, using as λ one of these greater values would imply excluding a even higher number of predictors. Given that our goal is to identify a common set of predictors between the 4 responses, it is better to keep a higher number of predictors in this step.

The following step basically consisted in finding the already mentioned intersection between the four sets of predictors selected for the four corresponding dependent variables. Finally, we used the same set of predictors to do a logistic regression for each of the response variables.

In the following subsection, we assess the meaning of the results obtained with the procedure here described.

4.3 Results

Firstly we performed the Lasso considering the richest decile in 2019 as the response variable.

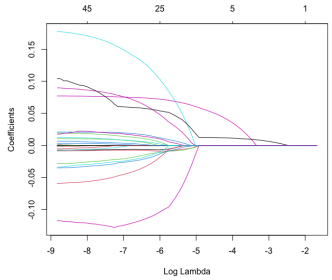

Figure 1:
The figure shows how coefficient estimates shrink with penalization

This plot is an example of a logarithmic transformation of λ in relation to regression coefficient magnitude, where each line represents a predictor variable. As λ (to facilitate interpretation we actually applied a logarithmic (log) transformation to λ) increases in size, we can notice that some of the regression coefficients start moving towards zero.

Subsequently, in order to select an appropriate tuning parameter, we performed the aforementioned k-fold cross validation. The effects of different λ values on the model accuracy metrics, that in this case are mean squared errors, are shown in the following chart:

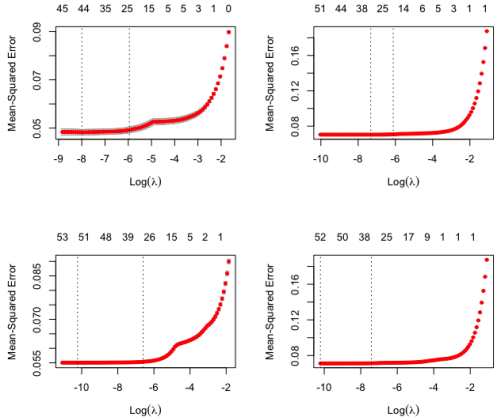

Figure 2:
Cross Validated MSE of Lasso fit for the different four response variable

These illustrate the sweet spots in terms of tuning the λ to be its optimal value. In the graph

in top left, that was made considering the richest decile in 2019 as response variable, the value at which λ results in the lowest amount of MSE is 0.0003372 (the value of the logarithm is around -8). However we decided to choose the value such that the logarithm is approximately equal to -6 (optimal λ + one standard deviation). Therefore the non-zero variables remain 25 unlike the initial 45. As concerning the second response variable, the richest quartile in 2019, the results suggests to choose a lambda equal to 0.0022068. The 32 starting variables are reduced to 20. Then the poorest decile in 2019 was considered as the response variable and choose λ equal to 0.0013627 leads us to remove 21 variables.

We then rebuilt the model using the previously found tuning parameters. In the following figure we considered only the first response variable (the richest decile in 2019). Each line represents a

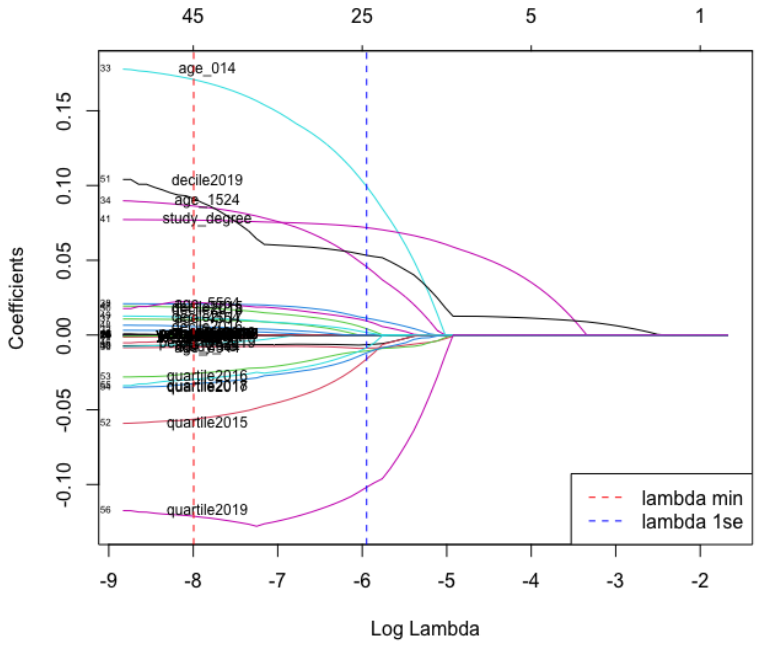

Figure 3: The figure shows how coefficient estimates shrink with penalization

predictor. The red dashed line indicates the minimum λ value but we focused on the blue dashed line that is one standard error distant. The chosen lambda value was, indeed 0.0026107 and the logarithm of the latter is about -6. Moreover, the upper part of the chart shows the total number of non-zero coefficients.
Among the most significant regressors, in this case, there seem to be the following: *age_014*, deciles from 2019, *age_1524*, study degree but also quartiles from 2019 to 2015.

We repeated the procedure with the other three response variables. If we consider the richest quartile in 2019, some of the most significant predictors are: quartiles from 2019, *age_014*, *age_1524* and study degree. The variables are the same with the poorest decile in 2019 as response variable. As concerning the poorest quartile in 2019 we found: *age_014*, *age_1524*, study degree, deciles from 2019, quartiles from 2019.
Regarding the interpretation of the prediction scores in terms of variance explained, it is interesting to observe that there is quite variability among the 4 different response variables and corresponding models. Indeed, in the first case the share of variance explained, measured by the Rsquare, is around 46%. This means that the model correctly predicts the response variable (belonging to richest decile in 2019 or not) in almost half the cases. The best model results the one with richest quartile as response variable, while there is a higher share of "failure" in the models predicting the poorest quartile and deciles.

Once all the predictors were determined by repeating the application of the LASSO, we tried to identify a unique core of significant variables for our analysis in all four cases. The resulting strong variables are 11:
Citizenship; *ripart4*, corresponding to the categorical variable of the geographical area of residence; Study title; *Y_dispo15*, which is the disposable income in 2015; being in the Northwest; being in the Northeast; *age_014*, belonging to the youngest class age (0 -14 years old); percentiles from 2019; deciles from 2019.
Identifying this common set of predictors aims at assessing the drivers of income level, regardless of the specific response variable (whether it corresponds to the richest or poorest part of the distribution). On the other side, it would be interesting as well, to assess which are the best predictors only for a specific section of the distribution and thus to focus only on the results for one specific response variable.

The results of these regressions reported in table 5 show that not all the predictors inserted in the regression actually prove to be statistically significant. Besides, there is some variability across the different regressions, corresponding to the different response variables: indeed, some predictors appear to be significant only for one dependent variable, e.g. the study title variables.

Table 5: Logit regressions output, where the independent variables derive from the common set of predictors from LASSO

	Dependent variable:			
	y_var1	y_var2	y_var3	y_var4
	(1)	(2)	(3)	(4)
cittad2	0.007*	−0.013***	−0.008*	−0.001
	(0.004)	(0.005)	(0.004)	(0.005)
cittad3	0.012***	0.006*	−0.003	0.009***
	(0.003)	(0.004)	(0.003)	(0.004)
ripart42	−0.001	−0.004**	−0.004***	−0.002
	(0.001)	(0.002)	(0.002)	(0.002)
ripart43	−0.003**	−0.001	0.001	0.003*
	(0.002)	(0.002)	(0.002)	(0.002)
ripart44	0.001	0.020***	0.015***	0.030***
	(0.001)	(0.002)	(0.001)	(0.002)
titostud1	−491,586,981.000	−256,954,200.000	−947,832,086.000*	192,808,827.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
titostud2	−491,586,981.000	−256,954,200.000	−947,832,086.000*	192,808,827.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
titostud3	−491,586,981.000	−256,954,200.000	−947,832,086.000*	192,808,827.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
titostud4	−491,586,981.000	−256,954,200.000	−947,832,086.000*	192,808,827.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
titostud5	−491,586,981.000	−256,954,200.000	−947,832,086.000*	192,808,827.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
titostud6	−491,586,981.000	−256,954,200.000	−947,832,086.000*	192,808,827.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
Y_dispo15	0.00001***	0.00000***	0.00000***	0.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
ye3_lav_aut19	0.00000***	0.00000***	0.00000	0.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
north_west				
north_east				
age_014	−491,586,981.000	−256,954,200.000	−947,832,086.000*	192,808,827.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
age_3544	−0.029***	−0.026***	−0.001	−0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
percentile2019	−0.005***	0.013***	−0.002***	−0.018***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
decile2019	0.084***	−0.026***	−0.048***	0.054***
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	491,586,981.000	256,954,200.000	947,832,087.000*	−192,808,826.000
	(515,139,176.000)	(607,400,301.000)	(530,667,363.000)	(603,570,989.000)
Observations	214,383	214,383	214,383	214,383
Log Likelihood	2,679.517	−32,640.280	−3,687.289	−31,284.440
Akaike Inf. Crit.	−5,323.034	65,316.570	7,410.579	62,604.890
Note:				*p<0.1; **p<0.05; ***p<0.01

5 Conclusion

In this research, we tried to investigate the income mobility in Italy, between the years 2015 and 2019. After a brief introduction to the topic and a summary of the recent literature developed around it, we described our dataset. The variables included regard the income distribution (?? autonomous, dependent, retirement): we created new variables to obtain the quartiles, deciles and percentiles of 2015 and 2019. Besides, we converted these variables in dummies (= 1 if the unit belongs to that percentile/decile/quartile, 0 otherwise). We also created dummy variables corresponding to the different class ages, study titles and geographical residence.

After this pre-processing actions, we started investigating the income mobility through multinomial and logit regressions as well as with transition probability matrixes. In Section 3, indeed, we first detailed the multinomial logistic regression, used to predict the transition probability of a certain observational unit, with a certain gender, degree and geographical residence, who belonged to a certain income quartile in 2015.

The main conclusions derived from this analysis prove that men have greater chances of transitioning to a different quartile during the 5 years accounted. The easiest transition, or better the highest proportion of women transitioning from one quartile to another, can be found between the first and second quartiles (the two poorest).

Secondly, we focused on logistic regressions. Again, it has been proved that women belonging to a lower decile in 2015 are less likely than men to move in the highest decile in 2019. Besides, it is more probable to improve the income positioning in the distribution when having an age between 15 and 54 than with an age greater than 54 years old.

In Section 4, we applied the LASSO to do feature selection. Since LASSO requires identifying a response variable to select the corresponding best predictors, we identified 4 different responses (richest and poorest decile in 2019, richest and poorest quartile in 2019).

We performed the LASSO and tested it with cross-validation, to choose the best value of the tuning parameter λ, which eventually was identified with the *1se* one. Given this value, we selected the best predictors for each response variable and consequently identified the core of common predictors among the 4 sets. Given this set of regressors, we replied the logit regressions, using as dependent variables the 4 responses mentioned above. Not all the regressors resulted statistically significant, but for sure the LASSO procedure was helpful in identifying the features more useful to predict the responses.

Given the analyses carried on in this paper, we can assess that the main drivers of income in Italy (in the period take into account) are age, specifically younger class ages, the geographical residence and the owning of a study title.

As for income mobility, the main drivers are gender, age, study title and geographical residence. Besides, it emerges that the categories that have the most probability of moving from one level of income to another are men, especially if belonging to younger class ages.

Finally, this work tried to contribute to the literature assessing the drivers of income mobility in

Italy. We focused on the period between 2015 and 2019: however, given the recent economic global crisis, related both to the pandemic and to the war in Ukraine, it would be interesting to carry on similar analyses considering years after 2019, to see how income mobility changed in these last times.

6 Appendix

Table 6: Regressions analysis output using the relative delta percentiles or deciles

	Dependent variable:	
	δ percentile	δ decile
	(1)	(2)
Gender	-0.536^{***}	-0.170^{***}
	(0.017)	(0.003)
Age 0-14	3.386^{***}	1.122^{***}
	(0.170)	(0.033)
Age 15-24	3.914^{***}	1.399^{***}
	(0.155)	(0.030)
Age 25-34	3.481^{***}	1.204^{***}
	(0.154)	(0.030)
Age 35-44	3.544^{***}	1.157^{***}
	(0.154)	(0.030)
Age 45-54	3.364^{***}	1.110^{***}
	(0.154)	(0.030)
Age 65-74	3.312^{***}	1.127^{***}
	(0.153)	(0.030)
Age >75	3.361^{***}	1.155^{***}
	(0.152)	(0.030)
Study title	0.933^{***}	0.243^{***}
	(0.024)	(0.005)
No study title	0.107^{**}	0.044^{***}
	(0.043)	(0.008)
Self-employment share	0.059^{**}	-0.048^{***}
	(0.024)	(0.005)
Employment share	-0.130^{***}	0.044^{***}
	(0.024)	(0.005)
Pension share	0.124	-0.001
	(0.085)	(0.017)

	Dependent variable:	
	δ percentile	δ decile
	(1)	(2)
Region Lombardia	−0.011	0.007
	(0.033)	(0.006)
Region Trentino Alto Adige	0.159***	0.063***
	(0.042)	(0.008)
Region Veneto	−0.015	0.002
	(0.043)	(0.008)
Region Liguria	−0.002	0.007
	(0.049)	(0.009)
Region Emilia Romagna	−0.057	−0.015
	(0.051)	(0.010)
Region Friuli Venezia Giulia	0.012	0.023***
	(0.038)	(0.007)
Region Toscana	0.017	0.007
	(0.041)	(0.008)
Region Marche-Umbria	−0.111***	−0.022***
	(0.042)	(0.008)
Region Lazio	−0.248***	−0.056***
	(0.040)	(0.008)
Region Abruzzo-Molise	−0.342***	−0.110***
	(0.050)	(0.010)
Region Campania	−0.433***	−0.133***
	(0.040)	(0.008)
Region Puglia-Basilicata	−0.470***	−0.140***
	(0.038)	(0.007)
Region Calabria-Sicilia	−0.492***	−0.151***
	(0.035)	(0.007)
Region Sardegna	−0.367***	−0.111***
	(0.053)	(0.010)

	Dependent variable:	
	δ percentile	δ decile
	(1)	(2)
First decile 2015	5.006***	1.494***
	(0.036)	(0.007)
Second decile 2015 d2	0.946***	0.492***
	(0.034)	(0.007)
Third decile 2015	0.570***	0.243***
	(0.034)	(0.007)
Fourth decile 2015	0.410***	0.155***
	(0.033)	(0.006)
Fifth decile 2015	0.325***	0.100***
	(0.033)	(0.006)
Sixth decile 2015	0.254***	0.063***
	(0.033)	(0.006)
Seventh decile 2015	0.222***	0.052***
	(0.032)	(0.006)
Eighth decile 2015	0.171***	0.037***
	(0.032)	(0.006)
Constant	−3.420***	−1.212***
	(0.155)	(0.030)
Observations	214,102	214,102
R^2	0.131	0.264
Adjusted R^2	0.131	0.264
Residual Std. Error (df = 214066)	3.782	0.734
F Statistic (df = 35; 214066)	922.247***	2,194.763***
Note:	*p<0.1; **p<0.05; ***p<0.01	

References

Stephen Aldridge. *Social mobility: A discussion paper*. Performance and Innovation Unit, 2001.

Anthony B Atkinson and Salvatore Morelli. Chartbook of economic inequality. *ECINEQ WP*, 324, 2014.

Anthony B Atkinson, Thomas Piketty, and Emmanuel Saez. Top incomes in the long run of history. *Journal of economic literature*, 49(1):3–71, 2011.

Robert J Barro. Inequality and growth in a panel of countries. *Journal of economic growth*, pages 5–32, 2000.

Richard V Burkhauser, Brian Nolan, and Kenneth A Couch. Intragenerational inequality and intertemporal mobility. 2011.

Demetrio Guzzardi, Elisa Palagi, Andrea Roventini, and Alessandro Santoro. Reconstructing income inequality in italy: New evidence and tax policy implications from distributional national accounts. 2022.

Stephen P Jenkins and Philippe Van Kerm. Trends in income inequality, pro-poor income growth, and income mobility. *Oxford Economic Papers*, 58(3):531–548, 2006.

Dimitris Kirikos, Bernard Njindan Iyke, HO Sin-Yu, Nicholaos Dritsakis, Pavlos Stamatiou, Yu Hsing, Cameron J Gable, Shalendra Sharma, Marco Tronzano, Stephen Foreman, et al. Income inequality and growth: New insights from italy. 2017.

Simon Kuznets. Economic growth and income inequality. *The American economic review*, 45(1): 1–28, 1955.