Exploring Income Inequalities in Italy

Eugenio Dalpozzo

September - December 2024

Contents

1 Introduction

Income inequality remains a critical issue in Italy, with disparities driven by regional, gender, and educational factors. The historical economic divide between the industrialized North and the less developed South continues to be a key determinant of inequality, as evidenced by persistent differences in income, employment opportunities, and social mobility [1]. This North-South divide, rooted in historical and structural factors, has proven resistant to policy interventions aimed at fostering convergence [2].

Gender inequality is another pressing concern. Women in Italy face significant wage disparities, earning less than men across most industries and professions despite higher levels of educational attainment in certain fields [3]. This income gap is compounded by the underrepresentation of women in high-paying STEM (Science, Technology, Engineering, and Mathematics) careers, as well as occupational segregation and systemic discrimination [4].

Educational disparities further exacerbate income inequality. While higher education is associated with increased earnings, studies have shown that a bachelor's degree alone often does not result in significant income gains, emphasizing the importance of advanced degrees for economic mobility [5]. Additionally, access to quality education is unevenly distributed across regions, further entrenching existing inequalities [6].

Recent empirical research highlights the role of fiscal policies, labor market dynamics, and demographic changes in shaping income inequality in Italy [7]. These studies underscore the need for targeted interventions that address structural barriers and promote equitable economic growth.

This work tries to analyze income inequality in Italy between 2015 and 2021 using a comprehensive dataset from ISTAT. By examining factors such as education, age, gender, regional disparities, and citizenship, the analysis provides insights into the determinants of income disparities and their evolution over time. The findings aim to inform evidence-based policies that can mitigate inequality.

2 Dataset

2.1 Data Description

The dataset utilized in this study, provided by ISTAT, comprises a representative sample, extracted with a two-layer stratification with constant probability of selection, of the Italian population in 2021, consisting of approximately 2.3 million individual observations. The dataset includes a wide range of baseline demographic and socioeconomic characteristics, such as sex, region of residence, nationality, age, and education level. Furthermore, it encompasses detailed income variables from 2015 to 2021, which are disaggregated into categories such as disposable income, pre-tax income, income from dependent work, self-employment, and pensions. Each observation is also associated with a household identifier, allowing for the reconstruction of household-level income, and a weight coefficient that estimates the number of individuals each observation represents in the Italian population. Notably, the sum of these coefficients corresponds to the total population of Italy, enabling population-level analysis through appropriate weighting.

It is important to note that while the dataset provides a robust approximation of the income distribution in Italy, it is not exhaustive. All income data are derived from traced fiscal records, which exclude significant components such as income from shadow economy and specific welfare transfers and invalidity pensions. Additionally, decisive elements such as income from capital gains and financial activities are not fully captured, and this shortcoming must be carefully taken into consideration. All information in the dataset is extracted as of 2021. Despite these limitations, the dataset serves as a reliable foundation for analyzing income inequalities and related socioeconomic factors in the Italian context.

2.2 Preliminary Checks

Before conducting the analysis, several data manipulations were carried out to prepare the dataset. First, new variables were created to facilitate future analyses. The variable for sex was transformed into a numerical format, with 1 representing females and 0 representing males. Additionally, some demographic characteristics were clustered. For instance, the original division into four macro-regions (North-East, North-West, Centre, and South) was simplified into two macro-regions: North versus South and Centre. Similarly, the original eight categories of educational level were aggregated into two broader groups: low-educated (individuals with up to a high school diploma) and high-educated (those with at least a Bachelor's degree). Age classes were also recategorized into three groups: young (up to 25 years old), adults (26 to 64 years old), and elderly (65 years and above).

Furthermore, data for the Harmonised Consumer Price Index (HICP) from ISTAT was used to adjust nominal income into real terms. The data represents the yearly average of the index, using 2015 as the reference base. This

adjustment accounted for inflation registered between 2015 and 2021.[1]

Preliminary checks were performed to ensure data quality and validity. First, the dataset was examined for duplicates, and none were found. Next, a missing value analysis was conducted to investigate whether the absence of disposable income data was correlated with any baseline characteristics. This was done using two methods: analyzing the frequencies of missing disposable income values for each characteristic and performing a chi-square test for interdependence between baseline characteristics and missing income values for each year.

Both methods revealed a significant correlation between missing disposable income values and certain demographic groups. For example, missing income data was more prevalent among individuals living in the South of Italy, females, and less-educated individuals. This finding underscores the importance of accounting for these biases when interpreting the results of subsequent analyses.

3   Descriptive Analysis

3.1   Italian Context

To provide a general background on income trends in the Italian context, I begin with an analysis of the evolution of disposable income from 2015 to 2021, focusing on individuals and households. Disposable income values were computed at both the individual, weighted by the sampling weight to reflect population-level averages, and household level, where the average was computed by aggregating the income of all individuals belonging to the same household, as identified by the codfam variable. The analysis is presented in two stages: first using nominal income values, and then adjusting for inflation by computing real income values using the Harmonised Index for Consumer Prices (HICP).

Figure 1a and 1c show the average disposable income trends at the individual level. In nominal terms, there is a clear upward trend, with the income increasing steadily from approximately €17,254 in 2015 to €19,054 in 2021. When adjusted for inflation, the growth appears more subdued, with average real disposable income increasing from €17,254 in 2015 to €18,198 in 2021. This indicates that while nominal incomes grew, part of this growth was offset by inflation, reflecting relatively modest gains in purchasing power over the period.

Similarly, Figure 1b and 1d display trends at the household level. In nominal terms, the average household disposable income increased significantly, from €28,594 in 2015 to €36,917 in 2021. After accounting for inflation, the increase is less pronounced as expected by a significant inflation period.

These trends highlight a gradual improvement in income levels for Italians. However, the impact of inflation reveals that the real gains in purchasing power are smaller than what nominal figures might suggest. This analysis establishes the economic background in which inequality phenomena can be further ex-

---

[1]ISTAT. (2023). *Harmonised Index for Consumer Prices (HICP)*. Retrieved from `https://esploradati.istat.it/databrowser/#/en/dw/categories/IT1,Z0400PRI,1.0/PRI_HARCONEU/IT1,168_757_DF_DCSP_IPCATC2B2015_1,1.0`.

plored, offering insights into the dynamics of income distribution in Italy during this time.

(a) Individual Nominal Income

(b) Household Nominal Income
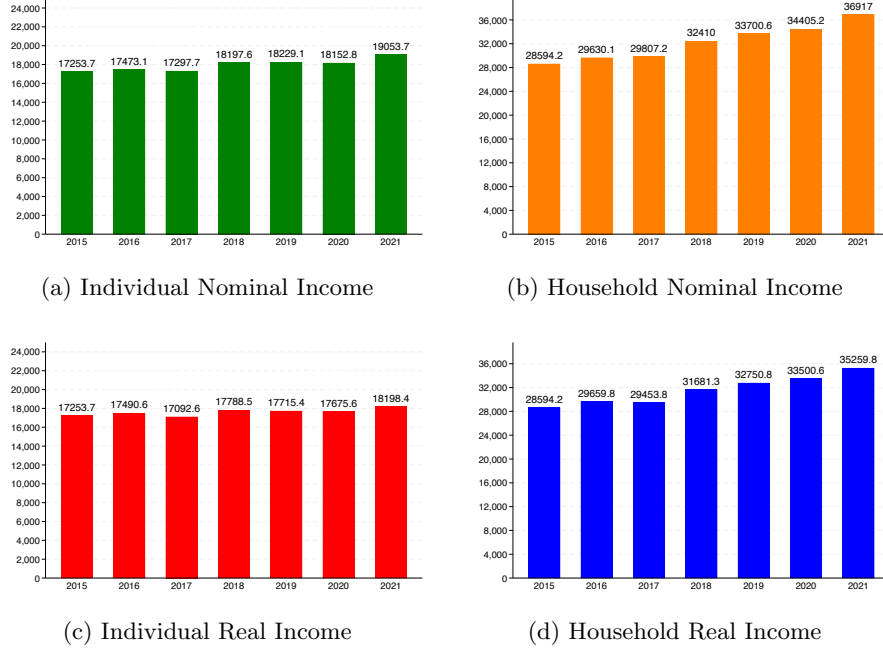
(c) Individual Real Income

(d) Household Real Income

Figure 1: Comparison of Nominal and Real Disposable Incomes for Individuals and Households (2015–2021).

3.2 Labour Market

Understanding the structure and evolution of the labour market is crucial for analyzing the broader economic context of Italy. Therefore, I categorized individuals into two groups for each year: those active in the labour market, and those retired. These categories provide valuable insights into the dynamics and trends of workforce participation and retirement, highlighting important socio-economic patterns.

To compute these variables, individuals were classified based on their income sources. Specifically, individuals active in the labour market were defined as those with positive employer income or self-employment income, excluding pension income. Retirees were identified as those receiving pension income exclusively, with no labor income sources. For each category and year, population-level values were computed using the sampling weight, and the sums for each category were calculated. Additionally, the yearly rate of change for each category between consecutive years was computed as a percentage change, offering a clearer view of trends over time.

Figure 2 illustrates the trends and rates of change for these categories from 2015 to 2021. The graph reveals that the number of retirees has steadily increased over time, reflecting demographic changes such as an aging population and longer life expectancies. In contrast, the population of those active in the labour market remains relatively stable across the years, indicating limited growth in workforce participation during this period. These findings emphasize the importance of monitoring the labour market's evolution to assess the pressing ageing problem.
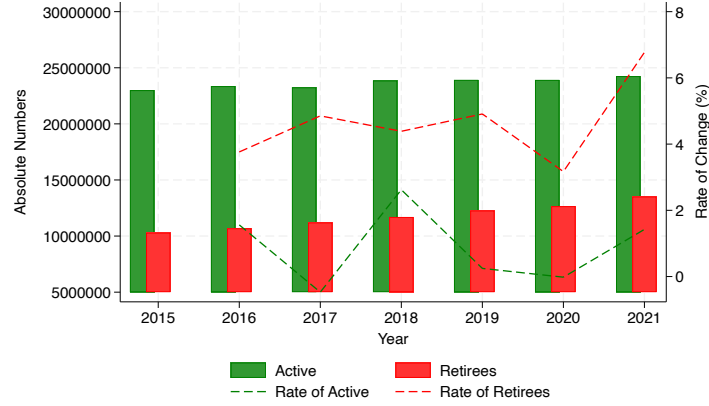
Figure 2: Labour Market Trends and Rates of Change in Italy (2015–2021).

3.3 Inequality Measures

Reliable measures of inequality are essential for providing accurate, data-driven statistics that can inform policymakers. These measures allow for a better understanding of the extent and nature of inequality, enabling the development of targeted policies to reduce disparities and promote social equity. By quantifying inequality, it becomes possible to identify the most affected groups, monitor changes over time, and assess the impact of interventions aimed at fostering a fairer distribution of resources.

3.3.1 Gini and Atkinson coefficients

Inequality measures such as the Gini coefficient and the Atkinson coefficient are fundamental tools for assessing the distribution of income within a population. These measures provide insights into the level of economic inequality, aiding policymakers in understanding disparities and implementing targeted interventions. The Gini coefficient, ranging from 0 to 1, quantifies inequality with 0 representing perfect equality and 1 indicating maximum inequality. The Atkinson coefficient, on the other hand, incorporates a degree of social aversion to

inequality, allowing for a more nuanced view of income distribution by emphasizing the lower end of the spectrum.

Table 1 presents the computed values of the Gini and Atkinson coefficients for the period 2015 to 2021. The table reveals a slight downward trend in inequality during this period, as indicated by both measures. The Gini coefficient decreased from 0.405 in 2015 to 0.395 in 2021, while the Atkinson coefficient showed a more pronounced decline, dropping from 0.368 to 0.332. This indicates some improvement in income equality over the years.

It is important to note, however, that these results are based on income data that exclude capital gains and financial income. On the other hand, undeclared labour income might regard more frequently the lower income deciles. These limitations can distort the analysis, as such income sources tend to be concentrated among the wealthiest individuals, thus contributing significantly to overall inequality. As a result, the measures presented here may underestimate the true extent of income disparities.

Year	Gini	Atkinson
2015	0.405	0.368
2016	0.404	0.365
2017	0.399	0.350
2018	0.400	0.351
2019	0.401	0.351
2020	0.400	0.340
2021	0.395	0.332

Table 1: Gini and Atkinson Coefficients for Italy (2015–2021).

3.3.2   Income Quintile Ratio

The income quintile ratio, often referred to as the S80/S20 ratio, is a widely used measure of income inequality. It compares the total income received by the richest 20% of the population (the 80th percentile and above) to the total income of the poorest 20% (the 20th percentile and below). This ratio provides an intuitive understanding of the disparity between the top and bottom income groups, making it an essential tool for analyzing income distribution and social equity.

Table 2 presents the computed S80/S20 ratios for Italy from 2015 to 2021. The table indicates a downward trend in the ratio over this period, suggesting a slight reduction in income inequality. For example, the ratio decreased from 3.855 in 2015 to 3.332 in 2021. This implies that the richest 20% of the population received approximately 3.3 times the income of the poorest 20% in 2021, compared to nearly 3.9 times in 2015.

These values are not fully consistent with official data from sources such as Eurostat. For example, Eurostat reports slightly different trends and values for

Italy's income quintile share ratio (S80/S20) during the same period[2]. Such discrepancies may result from the exclusion of certain income components, such as financial and capital incomes.

Year	S80/S20 Ratio
2015	3.855
2016	3.697
2017	3.662
2018	3.505
2019	3.601
2020	3.586
2021	3.332

Table 2: Income Quintile Share Ratio (S80/S20) for Italy (2015–2021).

3.3.3 Poverty Rates

Poverty rates are a crucial indicator of economic inequality and social well-being. They measure the proportion of households with a real income below 60% of the median income, a commonly accepted threshold for identifying those at risk of poverty. This metric provides policymakers and researchers with a valuable tool to assess the prevalence of economic hardship and the effectiveness of social welfare policies.

Table 3 displays the poverty rates for Italy from 2015 to 2021. The data reveals a gradual decline in poverty rates over this period, dropping from 21.3% in 2015 to 19.2% in 2021. This indicates a modest improvement in the economic conditions of lower-income households, which could be attributed to various factors such as social policies, economic growth, and labor market dynamics.

Unlike some other measures discussed in this document, the poverty rates presented here align closely with official statistics. For instance, data from sources like Statista confirms similar trends in Italy's at-risk-of-poverty rate[3]. This consistency enhances the reliability of the results and underscores the importance of addressing poverty as a central issue in public policy.

The most interesting categorization to explore was the differences in poverty rates across macro-areas of Italy. Figure 3 illustrates these trends, highlighting significant regional disparities. Over the years, poverty rates in the South consistently remained higher than in other regions, reaching levels over 30% in some years, while poverty rates in the North-East and North-West remained significantly lower, around 10–20%. The Centre of Italy lies between these extremes. This regional inequality underscores the persistent economic divide between Northern and Southern Italy and calls for targeted interventions to address structural disparities.

---

[2]https://tradingeconomics.com/italy/income-quintile-share-ratio-s80-s20-eurostat-data.html

[3]https://www.statista.com/statistics/619321/at-risk-of-poverty-rate-italy/

Year	Poverty Rate (%)
2015	21.319
2016	20.846
2017	20.800
2018	20.008
2019	20.223
2020	20.149
2021	19.219

Table 3: Poverty Rates in Italy (2015–2021): Households with Income Below 60% of Median Income.
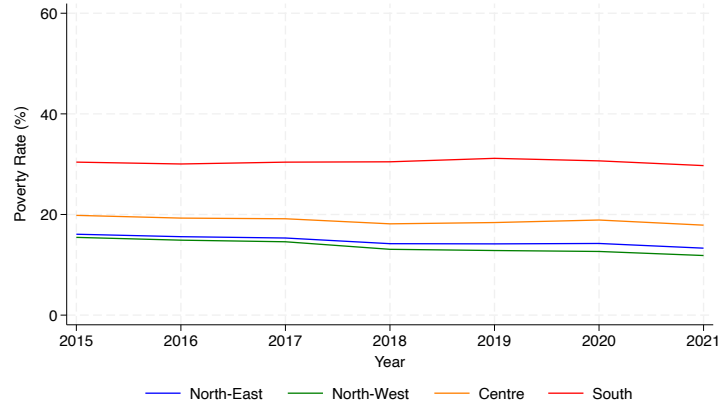

Figure 3: Poverty Rates by Macro-Area in Italy (2015–2021).

3.4 Gender Gap

The gender gap remains one of the most pressing source of inequality in contemporary societies, influencing various dimensions of life, including education, employment, and earnings. Addressing this disparity is crucial for achieving social equity and unlocking the full potential of the workforce. One critical aspect of the gender gap lies in the intersection of education and income, where women often face barriers that hinder their economic empowerment despite achieving higher educational qualifications in many cases.

Education plays a pivotal role in shaping opportunities and reducing inequalities. Interestingly, in higher education, there are actually more women than men obtaining advanced qualifications, as illustrated in Figure 4. This trend reflects the progress women have made in accessing education and excelling academically. For instance, the number of women with bachelor's, master's, and PhD degrees surpasses that of men in Italy.

However, this educational achievement does not translate into income equity. Figure 5 demonstrates the persistent disparity in average real individual income between men and women over time. While the income gap slightly narrows in some years, women consistently earn significantly less than men across all observed years (2015–2021). This disparity is a clear indication of systemic issues in the labor market, such as occupational segregation, wage discrimination, and the undervaluation of jobs traditionally performed by women.

The underrepresentation of women in high-paying STEM (Science, Technology, Engineering, and Mathematics) fields could be a potential explanation for the income gap. Encouraging more women to enroll in STEM education and pursue careers in these fields is vital for reducing the gender pay gap and ensuring a more balanced workforce. Addressing these disparities requires targeted policies and societal shifts to ensure that the progress women make in education is reflected in their economic outcomes.
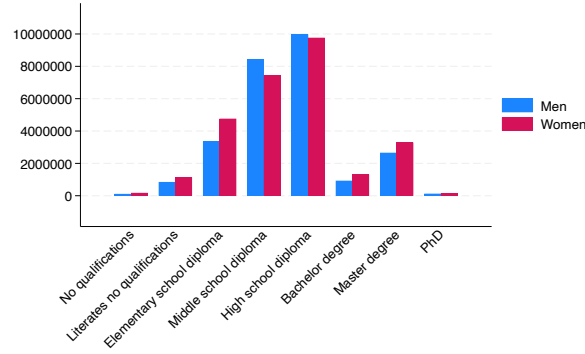

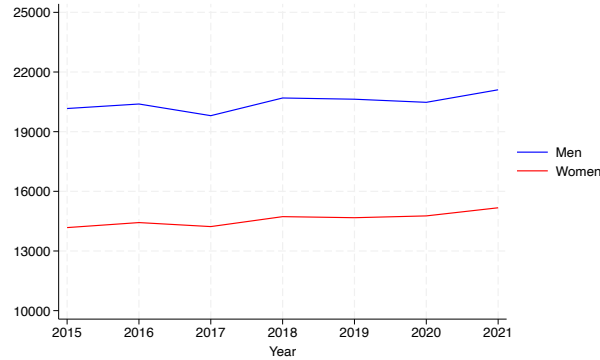Figure 4: Educational Attainment by Gender in Italy: Distribution Across Qualification Levels.

Figure 5: Real Individual Average Income by Gender in Italy (2015–2021).

3.5 Factors Determining Income Inequalities

Income differences in Italy are shaped by a variety of socio-economic and demographic factors. Understanding these differences is essential for designing targeted policies that address inequalities and promote a more equitable distribution of resources. This section explores key determinants of income disparities, analyzing trends based on education, age, place of residence, citizenship, and type of income. The following subsections provide a detailed discussion of these factors, supported by visual evidence.

3.5.1 Education

Education is one of the most significant determinants of income differences. Figure 6 illustrates the average income for individuals with varying levels of education from 2015 to 2021. As expected, individuals with higher education levels, such as master's and PhD degrees, earn significantly more than those with lower qualifications. However, an intriguing observation is that individuals with only a bachelor's degree earn, on average, incomes very similar to those with just a high school diploma.

This reinforces the common perception that, in many cases, holding a bachelor's degree is not sufficient to achieve high income levels. Instead, obtaining a master's degree or higher appears to be the key factor for significantly improving earning potential.

The persistent income gap between education levels underscores the importance of encouraging lifelong learning and ensuring equitable access to advanced education opportunities and it indicates structural barriers that limit upward mobility for those with lower education levels.
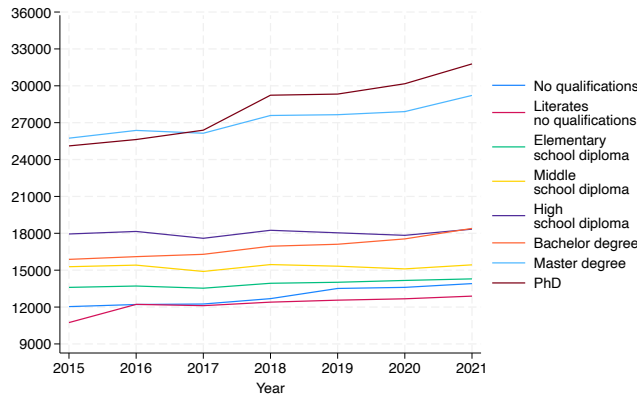

Figure 6: Average Income by Education Level in Italy (2015–2021).

3.5.2 Age

Age plays a crucial role in determining income, as shown in Figure 7. The highest average incomes are observed among individuals in the age range 45–74. Younger age groups (e.g. 15–24) and older individuals (75+) exhibit lower incomes, reflecting the life cycle effect and labor market participation trends. It is interesting to note a general positive trend over years in real incomes for younger generations aged 15–34. However, persistent differences highlight the importance of employment opportunities and social protection measures for the youngest and oldest cohorts.
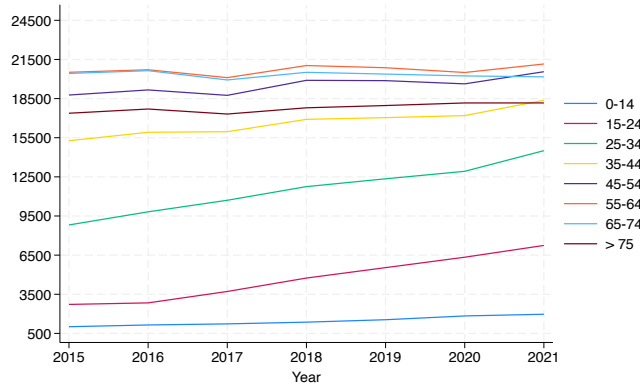

Figure 7: Average Income by Age Group in Italy (2015–2021).

3.5.3 Place of Residence

The historical economic divide between the North and South of Italy is one of the most well-documented disparities in the country. For decades, the North has enjoyed higher levels of industrialization, employment, and income, while the South has faced persistent challenges, including lower economic development, higher unemployment, and widespread poverty. These disparities are deeply rooted in historical, structural, and geographical factors and show no signs of being reversed.

Figure 8 highlights the regional disparities in income from 2015 to 2021. The North-East and North-West regions consistently exhibit the highest average incomes, followed by the Centre, while the South remains significantly behind. This pattern is stable across the observed period, indicating a persistent divide in economic well-being between the regions.

Despite numerous policies and interventions proposed to reduce these disparities, such as incentives for businesses to operate in the South and infrastructure investments, these efforts have proven largely ineffective. The data clearly show that regional income differences remain stable over time, underscoring the need for a more comprehensive approach to address this entrenched inequality.
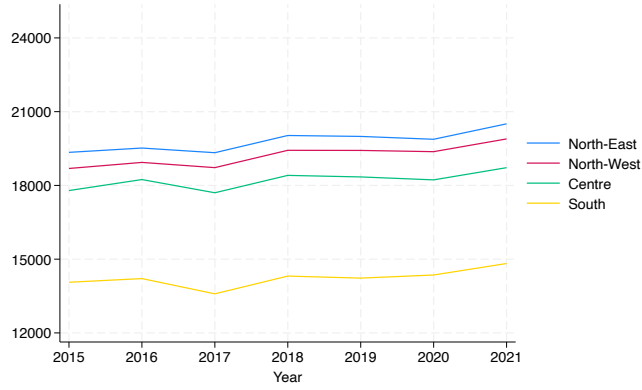
Figure 8: Average Income by Region in Italy (2015–2021).

3.5.4 Citizenship

Figure 9 reveals significant income differences based on citizenship. Italians earn the highest average incomes, while non-European Union (non-EU) citizens, particularly those from Africa and Asia, earn substantially less. These disparities reflect systemic challenges such as discrimination, access to the labor market, and differences in job opportunities for non-citizens.

The data could also suggest that much of the immigration to Italy consists of low-skilled individuals, often associated with lower educational levels and, consequently, lower incomes. This pattern highlights the structural issues in the labor market and the limited opportunities for immigrants to access higher-paying jobs or improve their economic standing.
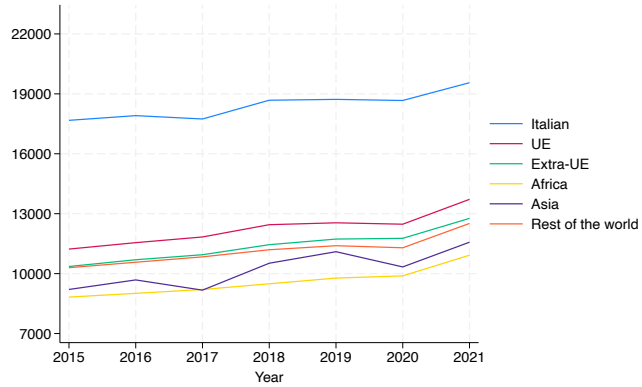

Figure 9: Average Income by Citizenship in Italy (2015–2021).

13

3.5.5 Type of Income

The source of income is another critical determinant of income differences. Figure 10 shows that income trends vary significantly across different categories: employees, retirees, and self-employed individuals. In the first three years of the analysis, the average income of self-employed individuals was surprisingly lower than that of employees and retirees. However, in recent years, the income of the self-employed has surpassed both employees and retirees, reflecting a notable shift.

This trend could suggest positive effects of anti-evasion policies implemented during the observed period, as the self-employed category is traditionally associated with higher levels of tax evasion in Italy. Increased compliance and transparency in reporting income may have contributed to this rise, narrowing the gap and ultimately surpassing other categories.

Despite these changes, retirees remain the group with the lowest average incomes, highlighting the need for continued attention to pension adequacy and financial security for the elderly population.
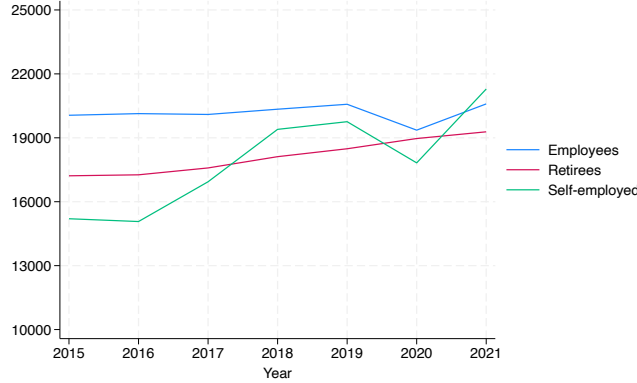

Figure 10: Average Income by Type of Income in Italy (2015–2021).

4 Inferential Analysis

4.1 LASSO method

To better understand the determinants of real disposable income in 2021, a LASSO (Least Absolute Shrinkage and Selection Operator) linear regression model was employed. This method is particularly useful for variable selection and regularization in models with multiple predictors, as it minimizes overfitting and improves prediction accuracy. The model included real disposable income from previous years (`y_dispoR20` to `y_dispoR15`) and categorical predictors for gender (`dummy_sesso_num1`), region of residence (`dummy_rip21`), education level (`dummy_titostud21`), and age group (`dummy_eta31` and `dummy_eta32`).

The LASSO model used cross-validation with 10 folds to identify the optimal penalty parameter (`lambda`). The selected `lambda` was 48.213, which included 10 predictors in the final model. The model achieved an out-of-sample R-squared of 0.7237, indicating that the predictors collectively explain approximately 72.37% of the variance in real disposable income for 2021. This high R-squared value highlights the strong predictive power of the selected variables.

The results of the analysis show that all the included predictors were retained in the final model, suggesting their importance in explaining income differences. The inclusion of past incomes (`y_dispoR20` to `y_dispoR15`) reflects the significant influence of income history on current income levels, while the categorical variables—gender, education, and age—indicate that demographic characteristics play a crucial role in determining income. The consistent selection of these variables underlines the importance of targeting disparities in these areas through data-driven policies.

4.2 Regression on disposable income

To investigate the determinants of real disposable income in 2021 (`y_dispoR21`), an extended regression model was estimated, including interaction terms between gender and region, education, and age groups. This model achieved an R^2 of 0.7334, indicating that approximately 73.34% of the variance in real disposable income is explained by the predictors. The addition of interaction terms provides deeper insights into how demographic factors interact to influence income levels.

Past income variables (`y_dispoR20` to `y_dispoR15`) are significant predictors, with coefficients for `y_dispoR20` (0.369) and `y_dispoR19` (0.420) being the largest. These results reaffirm the strong dependence of current income on historical income levels. Income from earlier years, such as `y_dispoR16`, has a diminishing influence, with its coefficient (-0.021) showing a small negative effect.

The inclusion of interaction terms reveals interesting dynamics. For gender and region (`dummy_sesso_num1 # dummy_rip21`), the interaction term (185.74) suggests that the income gap between genders differs across regions, with women in Northern regions experiencing higher incomes relative to men. Similarly, the interaction between gender and education shows a negative coefficient (-490.87), indicating that the gender income gap is wider among those with higher education levels, which highlights systemic inequities in labor market outcomes.

Age-related interactions also provide critical insights. The interaction term for gender and middle-aged individuals (`dummy_sesso_num1 # dummy_eta31`) is positive and substantial (1301.68), suggesting that middle-aged women earn significantly more than men in the same age group, controlling for other factors. Similarly, older women (`dummy_sesso_num1 # dummy_eta32`) also experience higher income advantages relative to men, with an interaction coefficient of 997.35.

The main effects of gender and region show significant disparities. For instance, the main effect of gender (`dummy_sesso_num1`) is negative (-205.08), sug-

gesting that women earn less than men on average, but this gap is moderated by regional and age-related factors. Region (`dummy_rip21`) also has a significant negative main effect (-102.50), reflecting lower incomes in Southern regions.

In summary, this extended regression model underscores the complex interplay between gender, region, education, and age in determining income levels. While past income remains the strongest predictor, the interaction terms highlight significant disparities and systemic inequities, suggesting the need for targeted policies to address these issues.

Table 4: Regression Results for Determinants of Real Disposable Income in 2021 (Partial Results)

VARIABLES	Coefficients	Robust Standard Errors
y_dispoR20	0.369***	(0.0875)
y_dispoR19	0.420***	(0.0749)
y_dispoR18	0.168*	(0.0871)
y_dispoR17	0.0574**	(0.0237)
y_dispoR16	-0.0210***	(0.00648)
y_dispoR15	0.0202	(0.0365)
1.dummy_sesso_num1	-205.1	(320.1)
1.dummy_rip21	-102.5	(70.57)
1.dummy_sesso_num1#1.dummy_rip21	185.7***	(60.58)
1.dummy_titostud21	-792.1***	(249.2)
1.dummy_sesso_num1#1.dummy_titostud21	-490.9***	(180.7)
1.dummy_eta31	1,061***	(188.3)
1.dummy_sesso_num1#1.dummy_eta31	1,302***	(147.5)
1.dummy_eta32	179.9***	(31.78)
1.dummy_sesso_num1#1.dummy_eta32	997.4***	(57.43)
Constant	883.4	(644.6)
Observations		1,348,589
R-squared		0.733

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.2.1 Diagnostic Checks

To evaluate the robustness and validity of the extended regression model, a series of diagnostic tests were conducted: variance inflation factor (VIF) analysis for multicollinearity, the Breusch–Pagan test for heteroskedasticity, and the Ramsey RESET test for omitted variable bias. The results are as follows.

The VIF analysis was performed to assess multicollinearity among the predictors, with a mean VIF of 4.08. Most predictors had VIF values below the conventional threshold of 10, indicating that multicollinearity is not a significant concern overall. However, `dummy_sesso_num1` exhibited a relatively high VIF of 9.61, suggesting a potential degree of multicollinearity. This might be attributed to the interaction terms involving this variable, such as `dummy_sesso_num1 # dummy_titostud21`, which itself has a VIF of 7.13. While these values approach concerning levels, they remain within an acceptable range, given the model's purpose and the theoretical importance of the interactions.

The Breusch–Pagan/Cook–Weisberg test for heteroskedasticity returned a chi-squared statistic of 5.80×10^8 with a p-value of 0.0000. This result strongly rejects the null hypothesis of constant variance, indicating the presence of heteroskedasticity in the model. Heteroskedasticity can lead to inefficient coefficient estimates and unreliable standard errors, suggesting the need to use robust standard errors to address this issue.

The Ramsey RESET test was employed to detect omitted variable bias by testing for nonlinear relationships in the model. The test returned an F-statistic of 2633.54 with a p-value of 0.0000, rejecting the null hypothesis that the model has no omitted variables. This indicates that the model may be underspecified, and additional nonlinear terms or omitted predictors might improve its fit and explanatory power.

In summary, while multicollinearity is generally not a major issue, the presence of heteroskedasticity and omitted variable bias suggests areas for improvement. Addressing these issues, such as using robust standard errors and considering additional predictors or nonlinear specifications, would enhance the reliability and robustness of the regression results.

4.3   Logistic Regression

This subsection presents the results of a logistic regression model aimed at analyzing the probability of being active in the labor market (`active21`) in 2021, as well as its determinants. The model includes variables reflecting prior labor market activity (`active20` to `active15`), demographic characteristics (gender, region, education level, and age groups), and interaction terms where appropriate. The regression achieves a Pseudo-R^2 of 0.6748, indicating that the model explains a substantial portion of the variability in labor market activity.

The odds ratios (`esttab, eform`) provide interpretable insights into the effects of each predictor on the likelihood of labor market activity. Past labor market activity is the strongest determinant, with `active20` having the largest odds ratio of 22.67. This suggests that individuals who were active in 2020 are over 22 times more likely to remain active in 2021 compared to those who were not active. The odds ratios for earlier years (`active19` to `active15`) progressively decrease, indicating that more recent activity has a greater influence on current labor market participation.

Among demographic factors, surprisingly gender (`dummy_sesso_num1`) shows an odds ratio of 1.239, indicating that females (coded as 1) are 23.9% more likely to be active in the labor market compared to males. Regional differences (`dummy_rip21`) have a modest but significant impact, with an odds ratio of 1.094, indicating slightly higher odds of labor market activity in Northern regions. Education level (`dummy_titostud21`), however, has a strong negative effect, with an odds ratio of 0.469. This suggests that individuals with higher education levels are less likely to be active in the labor market compared to those with lower education levels, possibly reflecting the longer duration spent in education.

Age is a critical determinant, as middle-aged individuals (`dummy_eta31`) and older individuals (`dummy_eta32`) exhibit much higher odds of being active com-

pared to younger groups. Middle-aged individuals are over 21 times more likely to be active (`dummy_eta31 = 21.17`), while older individuals have odds of 11.66 (`dummy_eta32`).

In conclusion, the results highlight the importance of past labor market activity and demographic characteristics in shaping labor market participation.

Table 5: Logistic Regression Results: Determinants of Labor Market Activity

VARIABLES	Odds Ratios	Robust Standard Errors
`active20`	22.67***	(0.0065753)
`active19`	3.081***	(0.0082700)
`active18`	1.786***	(0.0096228)
`active17`	1.357***	(0.0100772)
`active16`	1.274***	(0.0106558)
`active15`	1.276***	(0.0097882)
`dummy_sesso_num1`	1.239***	(0.0055026)
`dummy_rip21`	1.094***	(0.0055199)
`dummy_titostud21`	0.469***	(0.0077652)
`dummy_eta31`	21.17***	(0.0118696)
`dummy_eta32`	11.66***	(0.0111695)
Constant	0.009***	(0.0135112)
Observations	2,175,006	
Pseudo R-squared	0.6748	

Exponentiated coefficients (odds ratios); robust standard errors in parentheses. ***

$p < 0.001.$

5 Conclusions

This study has highlighted persistent income inequalities in Italy from 2015 to 2021, driven by structural factors such as the North-South divide and the gender pay disparities. Despite modest improvements in poverty rates and inequality measures, significant challenges remain.

Addressing income inequality is essential not only for social equity but also for economic growth and stability. Policies must target structural issues, such as improving access to education across regions, promoting gender equity in high-paying STEM fields, and fostering sustainable economic development in the South.

A data-driven approach is vital to designing effective interventions. Therefore, expanding datasets to capture omitted income sources, such as capital gains, could be useful to provide a more accurate picture of disparities.

In conclusion, reducing income inequality is both a socio-economic necessity and a pathway to a more prosperous Italy.

References

[1] Boeri, Tito, et al. "Regional Economic Disparities in Italy: A Persistent Divide." *Journal of Economic Perspectives*, vol. 29, no. 2, 2015, pp. 57-74.

[2] Trigilia, Carlo. "Development Policies and Regional Inequalities in Italy." *Cambridge Journal of Regions, Economy and Society*, vol. 11, no. 3, 2018, pp. 461-478.

[3] Mocetti, Sauro, and Giacomo Roma. "Gender Pay Gaps in Italy: Trends and Explanations." *Labour Economics*, vol. 42, 2016, pp. 67-85.

[4] Biagi, Federico, and Alessandra Chirico. "Women in STEM Fields in Italy: Challenges and Opportunities." *Italian Economic Journal*, vol. 6, no. 2, 2020, pp. 121-140.

[5] Checchi, Daniele. "Education, Inequality, and Income Distribution in Italy." *European Economic Review*, vol. 45, no. 5, 2001, pp. 913-930.

[6] Triventi, Moris. "The Role of Higher Education in the Italian Income Inequality Context." *Sociologia del Lavoro*, vol. 131, 2013, pp. 19-45.

[7] Jappelli, Tullio, et al. "Inequality and Redistribution in Italy: Evidence from Microdata." *Economic Policy*, vol. 35, no. 102, 2020, pp. 225-262.