Analyzing Income Inequalities across Italian regions: Instrumental Variable Panel Data, K-Means Clustering and Machine Learning Algorithms

Antonicelli, Margareth and Drago, Carlo and Costantiello, Alberto and Leogrande, Angelo

IULM University, Unicusano University, LUM University Giuseppe Degennaro, LUM University Giuseppe Degennaro

5 May 2025

Analyzing Income Inequalities across Italian regions: Instrumental Variable Panel Data, K-Means Clustering and Machine Learning Algorithms

Margareth Antonicelli

IULM University, Milan

margaret.antonicelli@iulm.it

Carlo Drago

Unicusano University, Rome

carlo.drago@unicusano.it

Alberto Costantiello

LUM University Giuseppe Degennaro, Casamassima

costantiello@lum.it

Angelo Leogrande

LUM University Giuseppe Degennaro, Casamassima

leogrande.cultore@lum.it

Abstract

This study examines income inequality across Italian regions by integrating instrumental variable panel data models, k-means clustering, and machine learning algorithms. Using econometric techniques, we address endogeneity and identify causal relationships influencing regional disparities. K-means clustering, optimized with the elbow method, classifies Italian regions based on income inequality patterns, while machine-learning models, including random forest, support vector machines, and decision tree regression, predict inequality trends and key determinants. Informal employment, temporary employment, and overeducation also play a major role in influencing inequality. Clustering results confirm a permanent North-South economic divide and the most disadvantaged regions are Campania, Calabria, and Sicily. Among the machine learning models, the highest income disparities prediction accuracy comes with the use of Random Forest Regression. The findings emphasize the necessity of education-focused and digitally based policies and reforms of the labor market in an effort to enhance economic convergence. The study portrays the use of a combination of econometric and machine learning methods in the analysis of regional disparities and proposes a solid framework of policy-making with the intention of curbing economic disparities in Italy.

1. Introduction

One of the most salient issues of economic analysis and public policy is the analysis of income disparities because income disparities influence social welfare, political stability, and economic development directly. Italy has a stunning example of economic disparities at the regional level with a lasting cleavage between the South and the North of the country. This economic cleavage has a long-lasting and structural character and it emerges through large disparities at the income and employment rates and investment in the infrastructure and development perspectives. The persistence of the disparities raises the crucial issue of the efficiency of economic policies in inducing the convergence of the different regions and the attainment of a lasting and sustainable economic development. An original methodological approach that combines cutting-edge econometric models and machine learning techniques can be applied in the analysis of economic disparities among the territories of Italy. The use of instrumental variable panel data models allows scholars to confront the issue of the problem of causality in the analysis of econometric models and the elimination of omitted variables' and endogeneity-induced biases. The employment of unsupervised methods of clustering, such as the K-Means algorithm with the application of the elbow criterion, facilitates the identification of homogeneous patterns of income distribution and the territorial categorization of territories based on the dynamics of inequality. The employment of machine learning models of income inequality prediction finally forms an invaluable quantitative analytical device for the identification of the major factors of economic disparity and the enhancement of the intervention abilities of public policies (Sbardella et al., 2021; Daniele, 2021; Guzzardi et al., 2024).

There exists a tremendous wealth of economic literature with comparative analysis of the disparities at the regional level and a wealth of development policy experiences with diversified results. Indonesia, for instance, has been affected by acute disparities at the regional level with the added COVID-19 economic crisis with results similar to those posted in Italy with the South of Italy experiencing a deeper economic recession compared with the North. In the US and Japan, agglomeration of wealth and expanding of the periphery exhibit similar patterns with Italy. Nevertheless, the development policies adopted in the two countries yield useful lessons with respect to the economic convergence strategies that may be adopted in the Italian experience. One of the strongest explanations of territorial disparities resides in public investment and infrastructure, that have been demonstrated powerful tools in the closure of territorial disparities. Between 1958 and 1986 a large public investment program effectively reduced Japanese territorial disparities. Italy has instead traditionally based welfare transfers instead of development strategies of a structural character and has therefore limited the effectiveness of economic cohesion policies. The experience of China teaches that selective investment in business incentives and infrastructure will stimulate a more equilibrated development among the economically disparate territories. One of the major factors in income inequality analysis is also access to energy sources and the role of digitalization. Recent empirical findings reveal that the diffusion of digital technologies has the power of bridging disparities and making economic opportunities in disadvantaged areas accessible. The existing North-South digital divide in Italy also constitutes a further impediment to economic convergence and the Mezzogiorno has lower rates of penetration of digital infrastructures and lower rates of digital skills acquisition. Furthermore, the transition of the energy sector has the power of substantially influencing the elimination of territorial disparities because the use of renewable sources and the

enhancement of energy efficiency immediately influence the cost of the production factor and the business competition of the area (Saeed and Siraj, 2024; Shudanko, 2024; Yang et al., 2020).

Regional disparities also need to be adjusted with the influence of social and institutional factors such as education, the state of the labor market, and institutional trust. The economic literature has widely established the correlation between economic inequality and the illiteracy index and has argued that an improvement in the provision of education in the least developed countries has the ability to lower income disparities. The state of the labor market with a large employment of temporary employment and a large sector of the working population with non-conventional forms of employment in the South also raises income disparities and diminishes social mobility. Higher disparities also connote lower institutional trust, a vicious cycle that has the ability of undermining the effect of development programs and inducing a cycle of suspicion and under-development. Methodologically, the study has an original framework with three pillars of analysis. To isolate the causal effect of the determinants of the disparities at the regional level, an instrumental variable econometric panel data model is used; this reduces the risk of the problem of endogeneity. To recognize patterns and homogeneous configurations of the Italian regions at the income disparities' level of analysis, a K-Means algorithm of the clustering technique is used. To recognize the best of the existing machine learning algorithm capable of correctly predicting the trends of the disparities of the income, the study uses a comparison of the models of the Support Vector Machines and Random Forest and the Decision Tree Regression (Giannoni et al., 2024; Daniele, 2021; Santamato et al., 2024).

The primary objective of this research is to provide a detailed diagnosis of regional inequalities in Italy, offering innovative quantitative tools for analyzing and predicting income distribution patterns. The findings of this study could offer concrete insights for designing economic policies aimed at reducing territorial disparities, suggesting specific interventions in areas such as education, technological innovation, labor policies, and infrastructure investment. To conclude, the analysis of the income disparities of the territories of Italy is not a quantitative comparison of cold statistics, but a vital aspect of the economic and social development problem of the nation. The results of the study put the need for an integrated approach that includes the application of econometric techniques, machine learning methods and evidence-informed economic interventions at the forefront. Only if we know the reasons and the mechanisms of the disparities in depth will it be possible to outline adequate policies that stimulate inclusive and durable growth in the entire national area (Guzzardi et al., 2024; Panzera and Postiglione, 2022; Asso, 2023).

The article continues as follows: the second section contains the literature review, the third section presents the variables of the model, the fourth section shows the methodology, the fifth section contains the data of net income inequality across Italian regions between 2004 and 2022, the sixth section shows the results of the k-Means clustering, the seventh section presents the econometric panel data model with instrumental variables, the eight section contains a confrontation among different machine learning algorithms, the ninth section explores the policy implications, the tenth section concludes.

2. Literature Review

2.1 Regional income disparity and economic inequalities

Indonesia constitutes a typical example of marked disparities of income at the regional level, compounded by the economic shock caused by the COVID-19 crisis. The crisis further sharpened pre-existing disparities and disproportionately hit the less developed peripheries and set in motion structural adjustments, especially in the labor market and sectoral structures. This development further entrenched the income disparities between the countryside and the cities and consolidated economic fragmentation at the territorial level. The same tendency unfolds in Italy with the COVID-19 crisis further sharpening the entrenched economic cleavage between the developed north and the structurally disadvantaged South. The Mezzogiorno has traditionally been marked by lower productivity and higher rates of employment and has seen a sharper downturn of economic activities compared with the north of the country. While Indonesia's regional disparities stem mainly from heterogeneities of infrastructure endowments and the distribution of resources, Italy's economic dualism has a foundation of long-standing structural disequilibrity based on the patterns of post-unification policy and convergent paths of industrial development. Both countries also demonstrate the occurrence of "convergence clubs," with the effect that particular regions experience economic catching-up and the rest of the countries fall behind and get stuck in low-income equilibria. Italy has a northward located high-income convergence club of the likes of Lombardy, Veneto, and Emilia-Romagna and a southern part with Calabria, Sicily, and Campania that has persistently lower income and limited prospects of economic ascendance (Akita and Alisjahban, 2023; Aginta et al., 2023).

America also forms an interesting comparison example. Gaubert et al., (2021) speak of a dual tendency toward the geographical agglomeration of economic wealth and the diffusion of poverty around the world. This tendency has specific relevance in Italy, with major cities in the north of the peninsula such as Milan and Bologna housing the wealthier classes and smaller villages and periphery areas of the South with economic stagnation. The "democratization of poverty" of the likes of Gaubert et al., (2021) captures the economic marginalization of southern Italy, with the limitations of the structural shortcomings of the diversification of the sector of the industry, the construction of the human capital and the public sector efficiency impeding economic development. The primary distinction between the US and Italy lies in the problem of the mobility of the workforce. The ease with which individuals within the US migrate internally and move toward the wealthier zones compared with the lower interregional mobility within Italy and the reinforcement of the persistence of the disparities between the zones.

There is a similar yet different example of how polarisation of ideology raises income disparities at the regional level that has been studied by Hyun-Chool and Repkine (2023). Korean government institutions and politics influence public investment and economic opportunity at the regional level and induce disparities in Korea's development. This has a similar parallel with Italy, with the latter's political inclinations at the regional level tending to drive the making of the government's policies and the general tendency of the North leaning toward business and the South toward state intervention and subsidies. Regional economic disparities in the two countries are associated with political forces and drive social cohesion and the government's programmes' effectiveness. South Korea's disparities also arise from rural and urban divisions and the agglomeration of industries, but Italy's disparities arise from stronger and deeper historical and institutional factors that preclude the prospects of convergence.

China's income disparities at the regional level, examined by Gao et al., (2022), take the form of a transition framework with economic liberalization and urbanization leading to dramatic contrasts between the coastal and interior regions. The situation has similarities with Italy's North-South contrast, with the latter's industrial and financial centers of the North witnessing steady growth and the South failing to keep up with it. The major distinction lies in the fact that China has had ambitious state-initiated

programs targeting the elimination of disparities at the regional level with the help of investment incentives and megaprojects of physical capital provision in the lagging areas. Italy has traditionally been dependent on subsidies and public sector employment as a source of subsidization of the South that has not been able to yield economic self-reliance in the long term. The record of the working of China's regional policies indicates that heavy investment in physical capital and connectivity has the potency of a powerful engine of reducing space disparities and Italy has not yet fully adopted this strategy.

Japan's historical experience with regional income inequality, as studied by Futagami (2022), provides another point of comparison. Between 1958 and 1986, Japan allocated substantial public investment to underdeveloped regions to reduce disparities, which led to significant convergence. This strategy contrasts with Italy, where regional policies have been less effective in fostering economic catch-up. While Japan successfully integrated peripheral regions into national economic growth through industrial policy and technological innovation, Italy's efforts have often been hindered by bureaucratic inefficiencies, rent-seeking behavior, and a lack of targeted investment in high-value industries. Japan's case demonstrates that long-term, strategic public investment can contribute to reducing regional disparities, a lesson that could be relevant for Italy's future regional development policies.

One of the explanations that tie with the disparities at the regional level involves the effect of institutional trust that has been explicated by Lipps and Schraff (2021) in the EU system. They argue that the persistence of economic disparities at the regional level leads to lower institutional trust and that this further entrenches economic stagnation. This ties specifically with the situation of Italy with the South traditionally having lower government institutions' trust relative to the North. The experience of ineffective government, corruption, and limited economic opportunity spawns a cycle of underdevelopment with skilled employees migrating to the wealthier zones and furthering the effect of the brain drain. This also occurs in the remaining EU countries with large disparities at the regional level such as Greece and Spain.

Aside from institutional variables, trade openness and foreign direct investment also influence regional income disparities, a factor that has been examined in the case of Sub-Saharan Africa by Xu et al., (2021). The study concludes that although trade liberalization increases economic growth in general, it favors developed regions at the expense of the underdeveloped ones, widening the disparities within a space. Italy's experience also fits this line of reasoning because globalization and the EU's integration efforts have consolidated the supremacy of the north's export-orientated firms at the expense of the South's inability to compete with the rest of the world. Encouragement of human capital and innovation in the lagging areas might help stem this uneven trade benefit distribution.

Finally, the case of Hungary's wine sector, analyzed by Ferto and Bojnec (2023), provides an industry-specific perspective on regional income inequality. They highlight how subsidies and market structures can influence income distribution within agricultural regions, a theme relevant to Italy, where agricultural subsidies under the EU's Common Agricultural Policy (CAP) have had mixed effects on regional development. While subsidies provide short-term income support, they do not necessarily lead to long-term structural improvements, a challenge that Italy's southern agricultural regions continue to face.

In conclusion, while regional income inequality is a widespread issue globally, Italy's case stands out due to its deep historical roots and structural complexity. Lessons from Indonesia, the United States, South Korea, China, Japan, and other regions offer valuable perspectives on potential policy interventions. Strategic investments in infrastructure, targeted industrial policies, improved governance, and greater labor mobility could contribute to narrowing Italy's North-South divide. However,

addressing long-standing historical and institutional barriers remains a formidable challenge, requiring coordinated efforts at both national and European levels.

2.2 Regional Disparities in Energy and Digitalization: Lessons from Global Studies for Italy

Income and energy access relate closely with each other, especially with regard to disparities at the regional level. The listed studies examine the ways in which energy poverty, digitalization, and the government affect economic disparities at the different world regions, specifically in the Asian and developing economies. Compared with the Italian experience, we find similarities and contrast regarding the ways the factors influence the regional disparities, specifically the entrenched North-South gradient. This analysis relates the findings of the listed studies with the experience of Italy with regard to the structural economic disparities, energy availability, and the transformation of the digital world as the major factors in the disparities at the regional level.

Acheampong et al., (2021) study how energy availability affects income inequality and contend that better availability of energy has the power of enabling the economically disadvantaged. This has particular application in the South of Italy, however, where energy facilities, while highly developed relative to the majority of developing economies, fall behind the north of the country in efficiency and sustain-ability and affordability. While the north of Italy has embraced higher use of the renewable and smart grid technologies, the South has a higher dependence on conventional sources of energy with higher expense and lower availability. This entrenches economic divisions with higher business operational expenses and lower business competitiveness. This contrasts with the situation in the majority of Asian developing economies where the issue of energy availability constitutes a primary economic inclusion impediment. The issue in Italy at the regional level is not necessarily availability of energy per se but the quality and sustain-ability of the distribution of energy.

Xu and Zhong (2023) study the correlation of income disparities-energy consumption with an emphasis placed on the moderating influence of digitalization. They conclude that the application of digital infrastructure has the likelihood of minimizing the pernicious influence of income disparities in the consumption of energy through the improvement of efficiency and availability. The north-south digital divide of Italy has a critical role in the widening of economic disparities. The large northward cities of Milan, Turin, and Bologna have highly developed digital infrastructure and judicious management of energy, while the southern half of the peninsula has a shortage of availability of broadband and shortage of digital skills in Calabria and Sicily. This prevents the application of energy-efficient technologies and reduces economic prospects. The findings of the study infer that investment in the development of digital infrastructure in behind-schedule areas has the likelihood of reducing the north-south energy gap and curbing income disparities.

Dong et al., (2022) study the reciprocal influence of income inequality, energy poverty and energy efficiency and whether it is the latter that leads to energy inefficiencies or the latter that leads to the former. This situation in Italy becomes clear in the way that inefficient energy consumption patterns drive economic disparities. The southern half of the nation has old and not so energy-efficient residential buildings and factories that drive higher energy spending relative to per capita income. Additionally, incentives such as home improvement subsidies and the installation of photovoltaic systems fall disproportionately among wealthier households that further drive existing disparities. This contradicts the situation in the majority of Asian and African developing economies where energy poverty relates to

the problem of basic access. The situation in Italy relates to disparities in the efficiency and afford-ability of energy consumption.

Acheampong et al., (2022) elaborate further the government's position in reducing energy poverty and the effect of political and institutional contexts toward the availability and afford-ability of energy. This has a direct application in the Italian scenario, wherein government inefficiencies and red tape impeded the application of energy transition strategies in the South. The north has the backing of superior institutions and private investment in green technology, whilst the South has been dependent on state intervention that has historically been ineffective based on corruption and management failures. The findings of the study estimate that superior systems of government, openness and focused investment will boost the availability of energy in the underdeveloped world and lower general inequality.

Tan and Uprasen (2021) specifically note the carbon-neutral future of ASEAN-5 countries and how income disparities affect the consumption of renewable energy. They conclude that the countries with higher disparities utilize lower investment in the consumption of renewables because wealthier societies can afford conventional sources of power and the poor do not possess the capabilities of turning toward that direction. The same situation happens with Italy with the north of the country leading the consumption of renewable power and the South lagging behind. The consumption of hydro and sun power has been successfully embedded in the Trentino-Aldo Adige and the Lombardy economies while Sicily and Calabria with the sun and wind power capabilities has been affected by the scale and finances barriers. This tells us that income disparities at the regional level need to be bridged if Italy has any hope of turning toward the consumption of renewable power and economic convergence will be fostered.

Li et al., (2022) explore the relationship between digital finance and sustainable development in China, highlighting how technological advancements can mitigate environmental and economic inequalities. In Italy, digital finance has seen significant growth, but disparities remain between northern financial hubs and southern regions with lower financial inclusion. The South's lower access to digital banking and investment platforms limits opportunities for small businesses and entrepreneurs, reinforcing income disparities. The study's findings suggest that improving digital financial literacy and accessibility in lagging regions could contribute to economic convergence.

Yin and Choi (2023) analyze the impact of digitalization on income inequality in G20 countries, concluding that technological advancements can help reduce disparities if widely accessible. This aligns with Italy's regional dynamics, where digitalization is a key factor in economic divergence. The South's slower adoption of digital technologies, partly due to lower educational attainment and infrastructure gaps, hinders economic growth and energy efficiency improvements. Addressing the digital divide through targeted investments in education and technology infrastructure could help reduce regional inequalities and support energy transitions.

Caragliu and Del Bo (2022) review the place of the smart city within the issue of urban inequality and the ways that technology has the power either to mitigate and widen disparities based on the availability and management of technology. Smart city projects in Italy have been focused in the north of the peninsula in cities such as Milan and Bologna with investment in AI-informed planning and green technology that enhances the standard of life and economic opportunity. Other southern cities fall behind the development of the smart technology that deprives them of the opportunity to offer higher-value business and talent. This further entrenches the larger cleavage of the South and the necessity of inclusive smart city strategies that place the development of the underdeveloped zones at the forefront of the solution.

2.3 Crime, Healthcare, and Economic Vulnerability: How Income Inequality Shapes Regional Development in Italy and Beyond

Income inequality has profound consequences for rates of crime, economic susceptibility, social cohesion, and access to healthcare, identified in numerous cross-country studies. Compared with the situation in Italy and the entrenched North-South cleavage there, we observe similar forces influencing regional disparities there as well. Italy has a developed economy with good social welfare provisions, yet the underlying disparities found there bear similarities with the patterns found in the emerging economies of Indonesia, Brazil, and China. Sugiharti et al. (2023) examine the correlation between income disparities and poverty and the criminality rates in Indonesia and conclude that the economic disparities also generate higher criminality rates. The correlation can be found in Italy with higher criminality rates in the poorer southern part of Italy consisting of the southern states of Campania, Calabria, and Sicily compared with the wealthier north of Italy. Organized criminality flourishes in the lower-income economic zones with higher rates of unemployment and restricted opportunity that drive individuals toward criminality. The scenario in Italy differs compared with Indonesia in the stronger institutional and historical foundation with the criminal organizations gaining long-term economic and political power within the impacted societies. Sun et al. (2022) also examine how the COVID-19 crisis further entrenched income disparities within the tourism sector that disproportionately affected southern Italy, the economy of which highly depends on the sector for output and employment. The sudden downturn of international travel severely affected cities such as Naples and Palermo and further amplified intra-region and region disparities. The findings align with general literature on economic resilience and sector dependency and highlight the need for southern Italy diversifying the economic base and the need for it to construct stronger and diversified industries.

Chen et al. (2021) emphasize that pandemics exacerbate global income inequality, a trend that was particularly evident in Italy. The COVID-19 crisis deepened the North-South economic divide, as southern regions faced greater challenges related to healthcare capacity, business closures, and labor market disruptions. Southern Italy's historically weaker healthcare infrastructure was further strained by the pandemic, exposing long-standing deficiencies. This pattern reflects the experiences of many developing countries, where pre-existing economic disparities were intensified, leaving vulnerable populations with limited resources for recovery. Kim et al. (2022) argue that social class and not income inequality in and of itself accounts for institutional trust. Institutional trust in Italy is lower in the south of the country and corruption and a failing government prevail there. Economic disparities certainly play a role, but lower citizen participation and deference of the rules also arise from a longstanding suspicion of the federal government. This aligns with findings in other countries in which the aggrieved citizenry, regardless of absolute income, has a pervasive suspicion of authorities and institutions of a democratic government. Sung et al., (2021) and Chi et al., (2024) note the effect of income disparities in the provision of the healthcare sector, specifically among the old population. Italy has been witnessing disparities in the quality of the provision of healthcare in the South and the North of the country with the north that is wealthy possessing superior facilities of the provision of healthcare, quicker waiting times, and specialist treatment. The South inhabitants migrate northwards in the quest for treatment, a situation similar with the wealthier part of China possessing superior facilities of the provision of the healthcare sector. The two works conclude that the improvement of the provision of the healthcare sector in the poor economic sections will help minimize the disparities. Sotomayor (2021) considers the use of minimum wage legislation in alleviating poverty and reducing inequality with the example of Brazil. Italy has no national

minimum wage and instead has sectoral collective agreements; yet wage differentials between the South and the North persist and southern workers receive lower compensation for similar employment. The application of a national minimum wage might help alleviate some of the disparities at the regional level, although economic reforms at the system-wide level would be required to tackle the underlying issues.

3. Data

Below we present the variables that have been used to investigate the inequality of net income with respect to the socio-economic and educational dimension. The data refer to the period between 2004-2022 and refer to the Italian regions. The data were acquired from ISTAT-BES. The variables are indicated in the following Table 1.

Table 1. Variables of the model.

Variables	Acronym	Definition
Net income inequality (S80/S20)	NII	Net income inequality (s80/s20): Ratio of total equivalent income received by the 20% of the population with the highest income to that received by the 20% of the population with the lowest income.
Inadequate literacy skills (students in third grade of lower secondary school)	ILS	Percentage of students in the third grade of lower secondary school who do not reach a sufficient level (Level I + Level II of 5 levels) of literacy competence.
Reading books and newspapers	RBN	Percentage of people aged 6 and over who have read at least four books a year (paper books, e-books, online books, audiobooks) for reasons not strictly scholastic or professional and/or have read newspapers (paper and/or online) at least three times a week out of the total number of people aged 6 and over.
Employed in fixed-term jobs for at least 5 years	ETF5	Percentage of fixed-term employees and collaborators who started their current job at least 5 years ago on the total of fixed-term employees and collaborators
Overeducated employees	OEE	Percentage of employed people who have a higher educational qualification than the most commonly held qualification for that profession out of the total employed people.
Irregular employees	IE	Percentage of employed people who do not comply with current labor, tax, and contribution regulations on the total number of employed people.
Employees who work from home	EWH	Percentage of employed people who have worked from home in the last 4 weeks out of the total employed people.

Source: ISTAT-BES. Link: https://www.istat.it/statistiche-per-temi/focus/benessere-e-sostenibilita/la-misurazione-del-benessere-bes/gli-indicatori-del-bes/ accessed 10/11/2024.

However, since we chose to apply the instrumental variable panel data we chose the following variables to use as instruments (Table 2).

Table 2. Instrumental Variables.

Definition	Acronym	Definition	Acronym
Internal Material Consumption	IMC	Beds in Residential Socio-Welfare and Socio-Health Facilities	BRSW
Urban Waste Produced	UWP	Elderly Treated in Integrated Home Care	EIDC

Urban Waste Disposal in Landfill	UWDL	Difficulty in Accessing Some Services	DAS
Contaminated Sites	CS	Irregularities in the Distribution of Water	IDW
Concern about Climate Change	CCC	Irregularities in the Electricity Service	IES
Satisfaction with the Environmental Situation	SEV	Seats-Km Offered by the Public Transport Service (Tpl)	SKM-TPL
Concern about the Loss of Biodiversity	CLB	Satisfaction with Mobility Services	SM
Research Intensity	RI	Assiduous Users of Public Transport	AUT
Propensity for Patenting	PP	Fixed Network Coverage for Ultra-Fast Internet Access	FNC-UF
Innovation of the Production System	IPS	Separate Waste Collection Service for Urban Waste	SWC
Knowledge Workers	KW	Beds for High-Care Specialties	BHCS
Cultural and Creative Employment	CCE	Hospital Emigration to Another Region	HER
Mobility of Italian Graduates (25-39 years)	MGI	Renunciation of Health Services	RHS
Regular Internet Users	RIU	General Practitioners with a Number of Patients Above the Threshold	GPP
Availability of at least one Computer and an Internet Connection in the Family	CICF	Doctors	DOC
Municipalities with Services for Families Entirely Online	MSFO	Nurses and Midwives	NM
Businesses with at least 10 Employees with Web Sales to End Customers	BWE10		

Source: ISTAT-BES. Link: https://www.istat.it/statistiche-per-temi/focus/benessere-e-sostenibilita/la-misurazione-del-benessere-bes/gli-indicatori-del-bes/ accessed 10/11/2024.
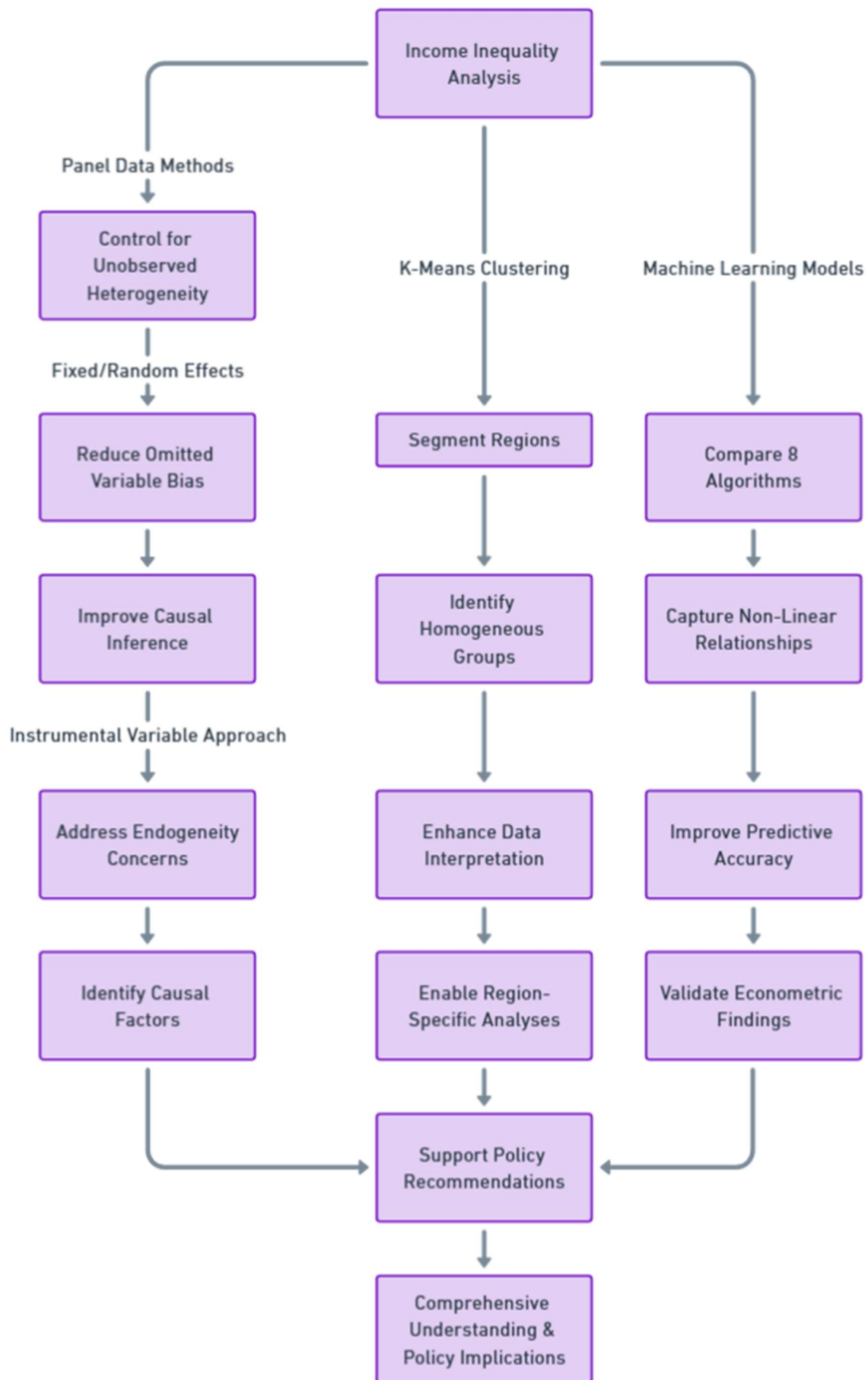
4. Methodology

By combining panel data instrumental variable analysis, clustering through k-Means, and a comparative assessment of eight different machine learning algorithms, this approach enhances both causal inference and predictive accuracy, offering a comprehensive understanding of income inequality dynamics. Panel data models provide a significant advantage in controlling for unobserved heterogeneity. Given that regional characteristics may include latent factors influencing income inequality, panel data methods allow for the estimation of fixed or random effects, thereby mitigating biases caused by omitted variables. This aspect is crucial when analyzing socio-economic phenomena where unobserved regional-specific attributes, such as cultural norms or historical economic structures, may play a role in shaping income distribution. The instrumental variable (IV) approach further strengthens causal inference by addressing endogeneity concerns. In socio-economic research, endogeneity can arise from omitted variable bias, measurement error, or simultaneity, where income inequality and certain explanatory variables influence each other reciprocally. By utilizing valid instruments—variables correlated with the endogenous regressors but uncorrelated with the error term—the IV methodology provides more reliable estimations of the causal effect of socio-economic and educational factors on income inequality. This methodological choice is particularly valuable in policy-oriented research, as it enables a more precise understanding of which factors genuinely drive disparities in income distribution (Huang, 2024; Munandar, 2023; Leogrande et al., 2023).

The incorporation of k-Means clustering adds an exploratory dimension to the analysis, facilitating the identification of homogeneous groups of regions based on socio-economic and educational characteristics. This method allows for a nuanced segmentation of Italian regions, revealing potential patterns that might not be immediately apparent through traditional econometric approaches. Clustering helps in classifying regions into distinct categories with shared characteristics, thus providing a richer interpretation of the data. For example, regions with similar levels of education, employment structure, and social mobility might exhibit comparable income inequality patterns, and clustering helps to empirically validate such hypotheses. Additionally, this segmentation can serve as a preliminary step for further econometric modeling by enabling differentiated analyses across regionally similar clusters. The comparative assessment of eight different machine learning algorithms enhances the robustness of the analysis by leveraging the predictive power of various modeling techniques. Machine learning algorithms excel in capturing complex, non-linear relationships among variables, an aspect that traditional econometric models might struggle to accommodate fully. By comparing multiple algorithms, the study ensures that the most effective model is selected based on predictive accuracy and generalizability. Some machine learning algorithms, such as decision trees, random forests, or gradient boosting methods, are particularly adept at identifying intricate interactions between socio-economic variables and income inequality. Moreover, machine learning techniques facilitate cross-validation, ensuring that the models generalize well beyond the sample data and are not merely overfitting to the specific dataset used in the analysis (Munandar, 2023; Kashwan and Velu, 2013).

This integrated approach—combining econometric rigor, clustering, and machine learning—provides a comprehensive and multifaceted understanding of income inequality across Italian regions. The synergy between traditional statistical techniques and modern data science methodologies allows for a more holistic investigation, balancing causal inference with predictive insights. Econometric methods, particularly panel data and IV models, contribute to a more rigorous examination of causal relationships, ensuring that findings are not merely driven by correlations but reflect genuine causal mechanisms. At the same time, clustering techniques enhance the interpretability of regional patterns, offering valuable insights into the socio-economic typologies of different areas. Machine learning models, with their ability to detect complex patterns and improve predictive accuracy, complement the econometric findings by providing an additional layer of validation and robustness (Figure 1).

Figure 1. Methodology.

```mermaid
graph TD
    A[Income Inequality Analysis]

    A -->|Panel Data Methods| B[Control for Unobserved Heterogeneity]
    A -->|K-Means Clustering| C[Segment Regions]
    A -->|Machine Learning Models| D[Compare 8 Algorithms]

    B -->|Fixed/Random Effects| E[Reduce Omitted Variable Bias]
    E --> F[Improve Causal Inference]
    F -->|Instrumental Variable Approach| G[Address Endogeneity Concerns]
    G --> H[Identify Causal Factors]

    C --> I[Identify Homogeneous Groups]
    I --> J[Enhance Data Interpretation]
    J --> K[Enable Region-Specific Analyses]

    D --> L[Capture Non-Linear Relationships]
    L --> M[Improve Predictive Accuracy]
    M --> N[Validate Econometric Findings]

    H --> O[Support Policy Recommendations]
    K --> O
    N --> O

    O --> P[Comprehensive Understanding & Policy Implications]
```

This strategy is particularly valuable for policy implications. Policymakers require reliable causal insights to design effective interventions aimed at reducing income inequality. The econometric component of this analysis provides these insights by distinguishing between factors that merely correlate with income disparities and those that actively drive inequality. Meanwhile, clustering allows for a tailored policy approach by recognizing that different regions may require distinct strategies depending on their socio-economic characteristics. Machine learning contributes by offering robust predictive tools that can be used for forecasting future inequality trends based on socio-economic shifts, thereby aiding proactive policy planning. By combining these three methodological pillars, the analysis achieves a balance between explanatory power, regional specificity, and predictive accuracy. Panel data and instrumental variables ensure a rigorous causal framework, clustering facilitates the identification of structurally similar regions, and machine learning enhances predictive capabilities and model validation. The integration of these approaches ultimately leads to a richer, more nuanced understanding of how socio-economic and educational factors influence income inequality across different Italian regions (Bruns-Smith et al., 2023; Condino, 2023).

4. Inequality of Income Across Italian Regions

Inequality across the Italian regions between 2004 and 2020. The analysis of net income inequality (S80/S20) across Italian regions from 2004 to 2020 reveals significant disparities in how income distribution has evolved over time. While some regions have successfully reduced inequality, others have experienced notable increases, particularly in the South, exacerbating the existing economic divide between northern and southern Italy. There is a slight fall in income inequality in Piemonte and Valle d'Aosta with absolute falls of -0.1 and a relative fall of -2.08% and -2.44%, respectively. This indicates a fairly stable economic scenario in the two regions with minimal deviation of the richest and the poorer part of the population. Trentino-Alto Adige and Emilia-Romagna also record slight falls in the S80/S20 ratio with a fall of -0.2 each. The falls imply that economic interventions or territorial dynamics stabilized the income distribution. Marche and Basilicata record the greatest falls in the S80/S20 ratio with falls of -9.76% and -12.24%, respectively. The improvements might be linked with social interventions, economic diversification, or demographic factors that contributed toward expanding the wealth distribution base. But a few of the areas registered notable increases in income disparities. Liguria registered an 11.54% improvement in S80/S20 ratio that signifies a widening of the wealth and poverty gulf. The 4.08%, 9.76%, and 7.14% respective percentage variation rates increases also take place in the areas of Lombardia, Veneto, and of the state of Lazio. The direction of the increases in the latter areas signifies that although they rank among the richest in Italy, the economic development may not necessarily be equally distributed. The same direction of the increases also follows in the states of the Friuli-Venezia Giulia and of the states of the state of the Umbria and of the state of the Molise, although the increases of the latter states are smaller in scale. The strongest trends follow in the southern half of Italy with the top- and lower-income brackets' gulf widening enormously. The highest 20.97% growth of the S80/S20 ratio has the state of the Campania with an estimate of 7.5 in 2020. This dramatic leap marks the entrenched economic deprivations of the state of the highest employment rates and limited economic opportunity. The large increases of 17.65% and 17.31% also follow in the states of the state of the Puglia and of the state of the Sardegna and the smaller increases of 4.92% and 4.35% also follow in the states of the state of the Calabria and of the state of the Sicily, respectively. The widening disparities of the latter four states reveal underlying structural factors such as lower-quality labor markets and lower rates of industrialization and entrenched socio-economic deprivations that disproportionately hit the

lower-income deciles. The evidence also highlights the widely-documented north-south economic cleavage of Italy. The north and center of Italy has experienced stable or slightly expanding disparities and southern Italy has experienced expanding disparities. Underpinning explanations of the trends will be the varied patterns of the development of the industry, employment prospects, government investment and social policies. Policy interventions against the disparities will be necessitated including investment in the education system, employment generation and development of the infrastructure with the intention of catalyzing the economic development of the South (Figure 2). Not adopting such interventions will further widen the disparities of southern Italy and entrench economic disparities at the expense of social mobility and economic prospects of future generations (Kocurová and Hampel, 2020; Daniele, 2021; Guzzardi et al., 2024; Rossi et al., 2024; Sbardella et al., 2021).

Figure 2. Inequality across the Italian regions between 2004 and 2020.

Inequality across the Italian macro-regions between 2004 and 2020. The net income disparities statistics of Italy's macro-regions during the period 2004-2020 demonstrate a widening and permanent economic distance between the South and the North. The income disparities have increased throughout the entire macro-region set except that the relative growth has been highest in the Mezzogiorno and the South

specifically. The income disparities growth has been slight in the North. The general S80/S20 ratio of the macro-region of the north has gone up from 4.7 in 2004 to 4.9 in 2020 with a 4.3% increase. Within this macro-region, the increment has been slightly higher at 6.1% with the ratio going up from 4.9 to 5.2. This signifies that economic growth in the industrious areas has been good and that wealth has been slightly unevenly distributed. The macro-region of the North-East with the territories specializing in powerful small and middle-size firms had a slight increment of 2.3%, and the ratio had gone up to 4.8-5.2. This signifies that income disparities in this macro-region had been higher and had not changed that dramatically compared with the macro-region of the North-West. The income disparities had gone up in the center of Italy from 4.8 to 5.2 with an 8.3% increment. This signifies that economic disparities had been widening in this macro-region that has the major cities of Rome and Florence among them. The economic disparities may be explained by the workforce transformation in the employment markets and the rural-urban disparities and economic reforms that had benefited the higher-income classes disproportionately. The highest and the most disquieting tendency has been the widening disparities of the Mezzogiorno macro-region with a general increment of the ratio of 6.0 in 2004 to 6.5 in 2020 with an 8.3% increment. In the Mezzogiorno, the South (Sud) posted the steepest growth at 6.4 and 5.7 and a 12.3% relative growth—the highest relative growth among the macro-areas. This reflects an widening gap between the richest and the poorest households in the South and possibly the effect of underlying economic structural issues and the influence of high rates of unemployment and limited economic prospects. The islands (Sicily and Sardinia) also posted a notable growth in the S80/S20 ratio with the latter advancing at 7.0 and an associated 6.1% growth. This further reflects economic hardships in the two regions with heavy dependence on seasonal activities such as agriculture and the tourism sector that leads to income volatility and economic disparities. Nationally, Italy's S80/S20 macro area also posted an increase at 5.9 and 5.6 and reflects a 5.4% growth in income disparities during the 16 years of observation. This development emphasizes the necessity of the efforts of the country toward equitable economic development. The north has been able to keep the problem of inequality at bay at least to a certain limit, the South and the islands posted widening disparities that reinforce Italy's old economic dualism. The statistics further imply that unless the government intervenes with specific interventions such as investment in the South's infrastructure and education and the provision of employment—these disparities between the two areas will be hard to stem and widen further in the future leading to further social and economic cleavages (Figure 3). The widening disparities in the center and north also imply the necessity of economic development interventions that encourage inclusive economic development and that the wealth that ensues will be dispersed among a larger population and not restricted among the higher-income earners (Sbardella et al., 2021; Culotta, 2021; Aresu et al., 2023).
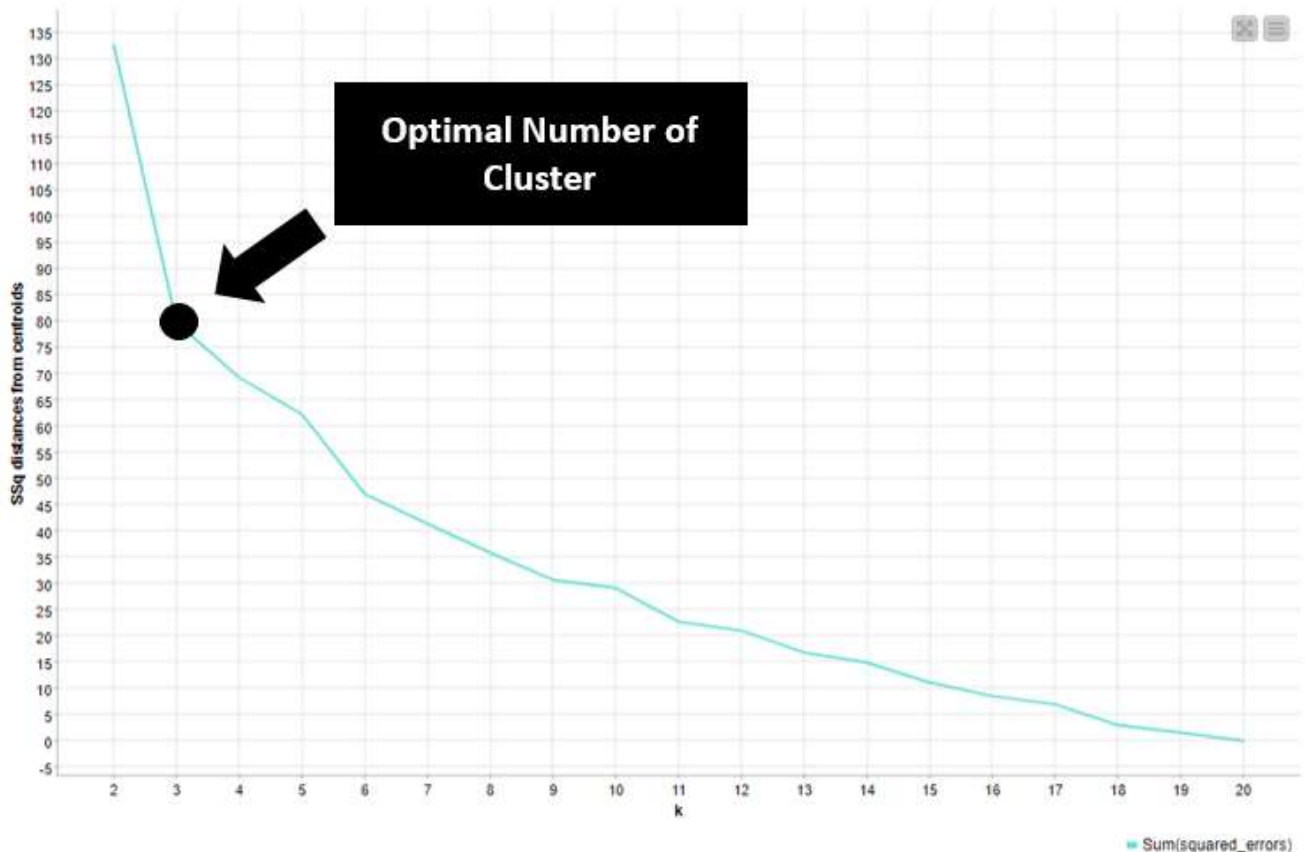
Figure 3. Inequality across the Italian macro-regions between 2004 and 2020.
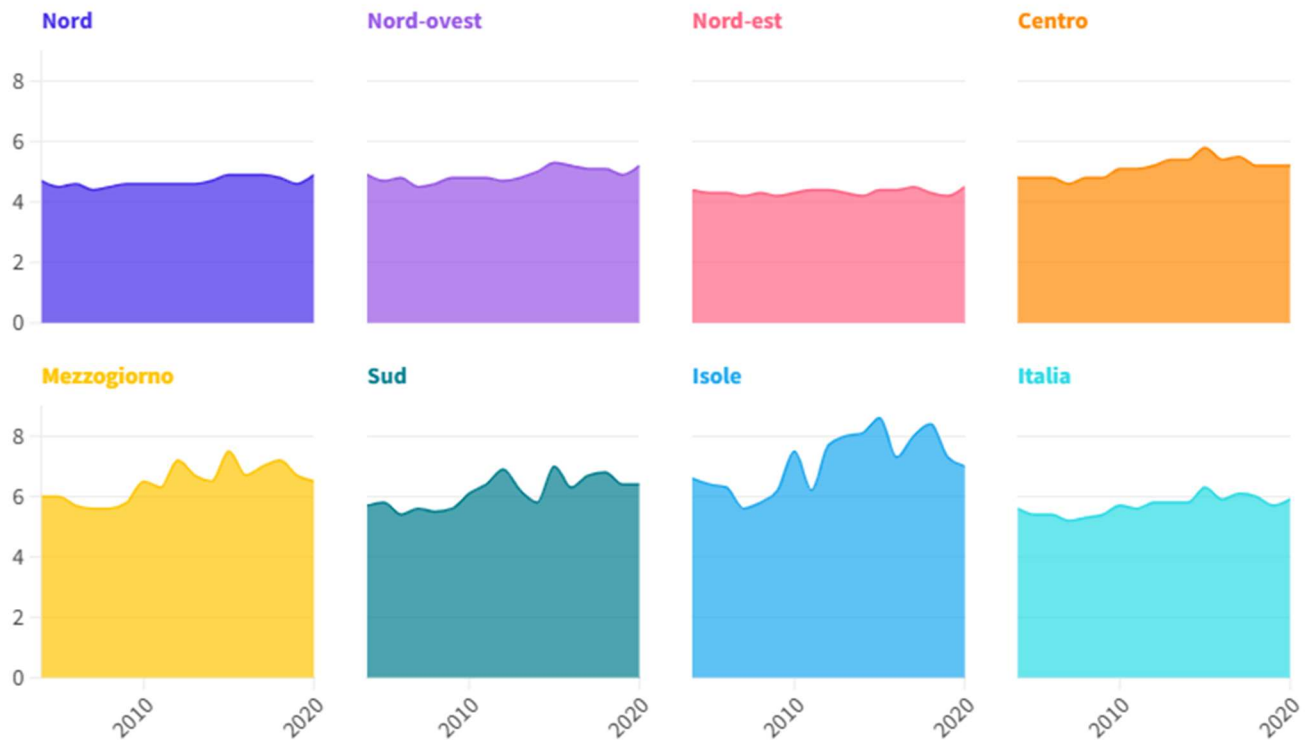
5. K-Means Clustering Optimized with the Elbow Method

Below we present a k-Means clustering algorithm optimized with the Elbow method. The optimal number of clusters with this method is k=3 (Figure 5).

Figure 5. K-Means Clustering optimized with the Elbow Method.

We have identified the three distinct clusters of the net income inequality index (S80/S20) of the Italian regions in 2020 through the application of the k-Means algorithm with the optimization of the Elbow method. The use of the clustering technique enables us to place the similar inequality regions into categories and identify patterns that will help policymakers and researchers better comprehend the disparities of the regions.

Cluster C1 includes the territories of the middle-to-high income disparities such as Liguria, Lombardia, Lazio, Puglia, Sardegna, Basilicata, Molise, and Piemonte. The S80/S20 values of the latter lie between 4.3 and 6.1 and suggest that income disparities do not reach the extremes of the most heterogeneous territories. The highest within-cluster inequality lies in the territories of Lazio, Liguria, Puglia, and Sardegna with a value of 6. The latter include large cities and economically relevant areas such as Rome and Milan, wherein agglomeration of wealth and the cost of life induce disparities. The values of 4.3–4.7 of Piemonte and Basilicata instead testify to a more equilibriated income distribution and might be explained by the latter territories' economic system and welfare interventions.

Cluster C2 comprises the lower-income-inequality areas of Valle d'Aosta, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, Toscana, Umbria, Marche, and Abruzzo. The S80/S20 rates of 3.7-4.7 of the regions demonstrate a lower income distribution compared with the national benchmark. Trentino-Alto Adige and Valle d'Aosta demonstrate the least inequality with 4.0 and 4.1 rates, perhaps due to the powerful regional economies and employment rates and good social programs. Emilia-Romagna and Veneto also fall within this category with the support of the highly developed cooperative and industrial sector that redistributes wealth. The inclusion of the likes of Toscana and Umbria signifies that the economic system of central Italy has a fairly stable balance regardless of the rural and the belt issues of the urban belt.

Cluster C3 also prioritizes the zones with the widest income disparities including Campania, Calabria, and Sicily. The latter zones have the widest disparities with S80/S20 values of 6 and above. The widest income disparities throughout Italy appear in Campania at 7.5, signaling entrenched economic hardships and higher rates of employment and underlying shortcomings of the labor market. Calabria and Sicily follow with 6.4 and 7.2 further confirming the entrenched economic cleavages between the north and the south of Italy. The latter zones historically had lower rates of industrialization and higher rates of dependence on the shadow economy and lower per capita income and thus larger income disparities.

The use of the technique of k-Means clustering results in a clear visualization of the geographical patterns of income inequality in Italy. The emergence of a high-inequality cluster (C3) in the South emphasizes the continuity of the latter territories' socio-economic problems, while the lower-inequality cluster (C2) encompasses the bulk of the North and the central peninsula territories. The mixed-profile cluster (C1) includes highly developed and economically diversified territories with the intensity of inequality conditioned by the economic patterns of the area. The findings emphasize the need for focused interventions that take into consideration the income disparities and particularly in the South economic reforms and investment may help bridge the gap of inequality and stimulate a more inclusive development (Figure 6).

Figure 6. Clusters.
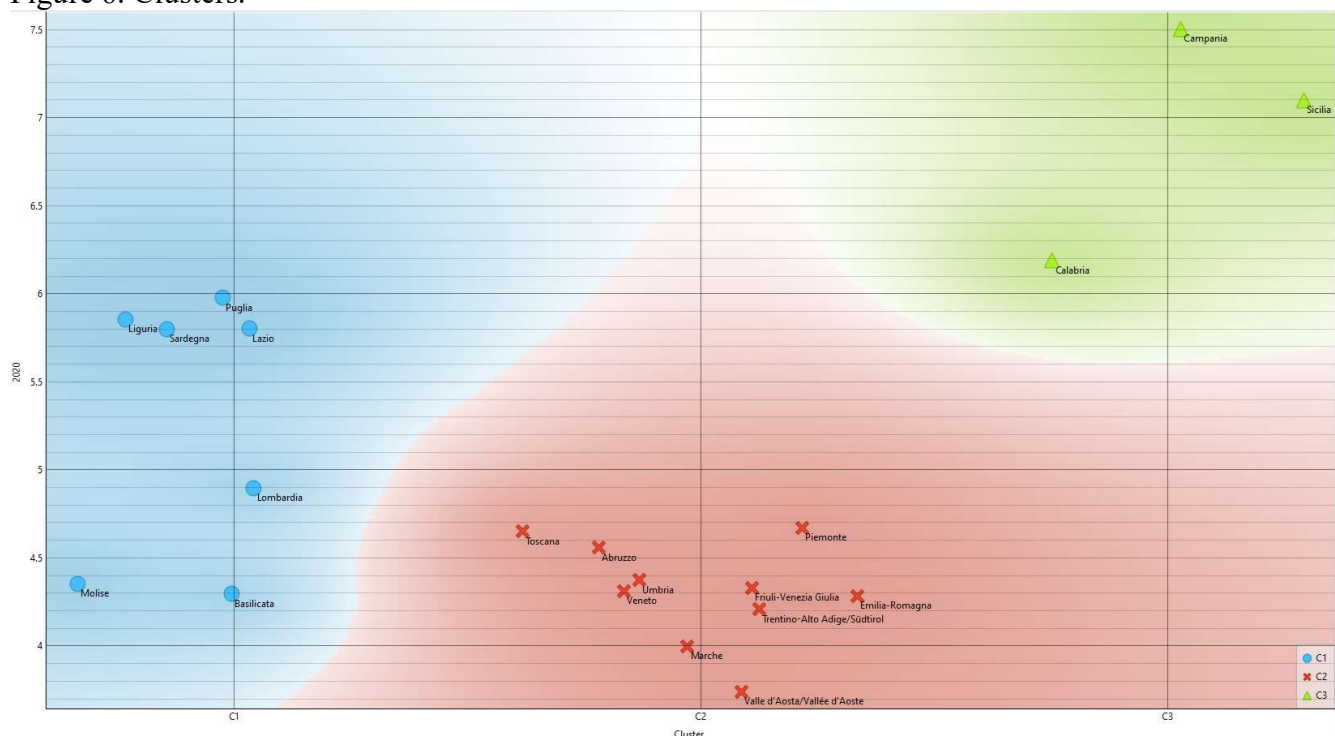
Median values of the clusters. The median values of net income inequality (S80/S20) for the three clusters identified through the k-Means algorithm highlight significant disparities between different groups of Italian regions. These median values provide a clearer understanding of the typical inequality levels within each cluster, reinforcing the economic divide between the North and the South. Cluster C1, with a median S80/S20 value of 5.3, represents regions with moderate inequality. This cluster includes a mix of economically strong areas such as Lombardia and Lazio, alongside regions with structural economic weaknesses like Puglia and Sardegna. The relatively high median suggests that, despite economic development, these regions still face significant income disparities. In areas with major urban centers,

such as Lazio (home to Rome) and Lombardia (home to Milan), wealth concentration and cost-of-living differences likely contribute to maintaining inequality at a moderate level. The presence of less developed regions like Puglia and Sardegna in the same cluster suggests that economic growth alone is not sufficient to guarantee a more equitable income distribution, as local labor market conditions and social policies play crucial roles (Korotaj et al., 2023; Tubadji et al., 2022).

Cluster C2, with a median of 4.3, represents regions with lower levels of inequality. This group includes many northern and central Italian regions, such as Trentino-Alto Adige, Veneto, Emilia-Romagna, and Toscana, which are known for their strong regional economies, well-developed industrial sectors, and effective social welfare systems. A lower median value suggests that these regions have a more equitable distribution of income, likely due to better employment opportunities, higher wages, and strong cooperative economic models that support the middle class. The presence of regions like Marche and Umbria, which traditionally have balanced economic structures, further supports this trend. The fact that the median value remains above 4.0 indicates that income disparities still exist, but they are far less pronounced than in other clusters. Cluster C3, with a median of 7.0, reflects the highest levels of inequality and includes southern regions such as Campania, Calabria, and Sicily. This cluster's high median indicates severe income disparities, where the richest 20% of the population earn at least seven times more than the poorest 20%. This extreme inequality is a clear sign of deep-rooted economic struggles, including high unemployment, a weaker industrial base, and reliance on informal economies. Campania, Calabria, and Sicily have historically faced socio-economic disadvantages, with lower investment levels, weaker infrastructure, and fewer job opportunities compared to the North. The persistence of such high inequality levels suggests that economic policies have not been sufficient to reduce disparities in these regions (Rossi et al., 2024; Guzzardi et al., 2024; Kocurová and Hampel, 2020).

Overall, the distribution of median values across the three clusters highlights Italy's ongoing economic divide. While northern and central regions maintain more balanced income distributions, southern regions continue to struggle with severe inequalities. These findings reinforce the need for targeted interventions to support economic development in the South, including investments in education, job creation, and infrastructure to create a more inclusive and equitable economy (Figure 7).

Figure 7. Median values of the clusters.
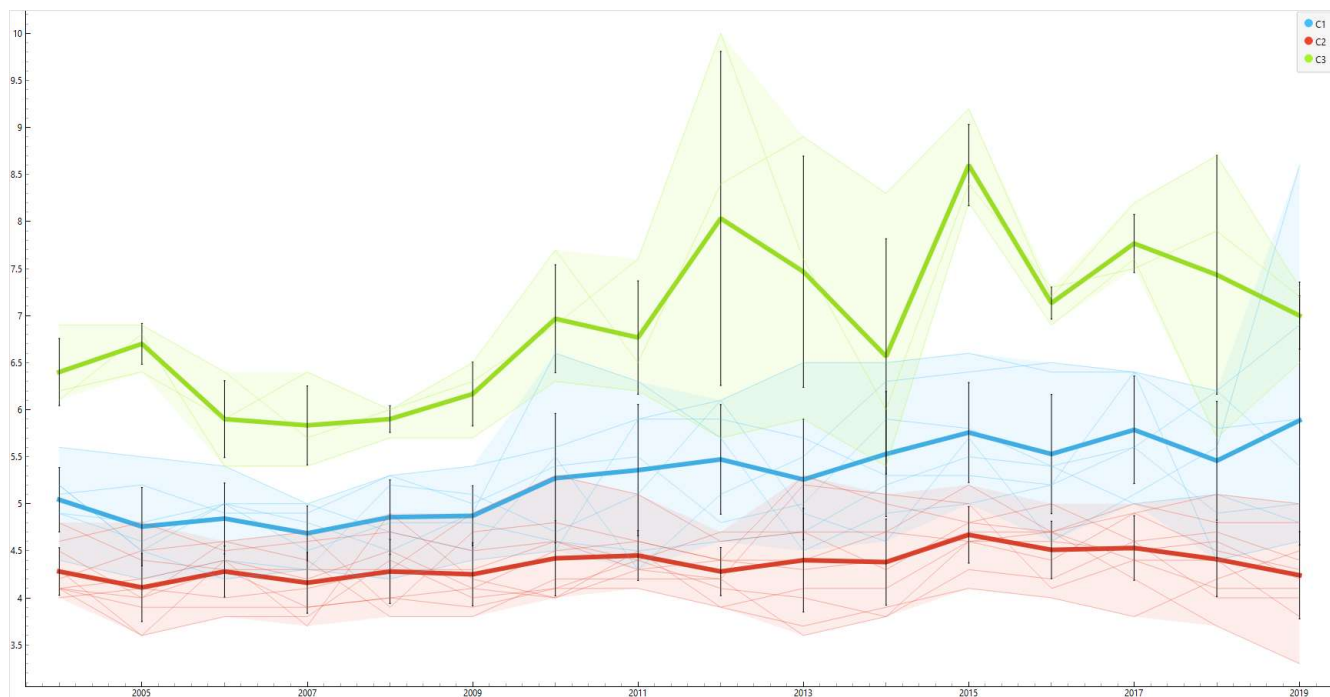
6. Instrumental Variable Panel Data to Estimate the socio-economic determinants of Net Income Inequality across Italian Regions

To estimate income inequality in the Italian regions we used panel data with instrumental variables as indicated in the formula below:

$$\begin{aligned}
NII_{it} = \gamma_0 &+ \gamma_1(ILS)_{it} + \gamma_2(RBN)_{it} + \gamma_3(EFT5)_{it} + \gamma_4(OEE)_{it} + \gamma_5(IE)_{it} + \gamma_6(EWH)_{it} \\
&+ \delta_1(AUT)_{it} + \delta_2(BHCS)_{it} + \delta_3(BRSW)_{it} + \delta_4(BWE10)_{it} + \delta_5(CCC)_{it} + \delta_6(CCE)_{it} \\
&+ \delta_7(CICF)_{it} + \delta_8(CLB)_{it} + \delta_9(CS)_{it} + \delta_{10}(DAS)_{it} + \delta_{11}(DOC)_{it} + \delta_{12}(EIDC)_{it} \\
&+ \delta_{13}(FNC - UF)_{it} + \delta_{14}(GPP)_{it} + \delta_{15}(HER)_{it} + \delta_{16}(IDW)_{it} + \delta_{17}(IES)_{it} \\
&+ \delta_{18}(IMC)_{it} + \delta_{19}(IPS)_{it} + \delta_{20}(KW)_{it} + \delta_{21}(MGI)_{it} + \delta_{22}(MSFO)_{it} + \delta_{23}(NM)_{it} \\
&+ \delta_{24}(PP)_{it} + \delta_{25}(RHS)_{it} + \delta_{26}(RI)_{it} + \delta_{27}(RIU)_{it} + \delta_{28}(SEV)_{it} \\
&+ \delta_{29}(SKM - TPL)_{it} + \delta_{30}(SM)_{it} + \delta_{31}(SWC)_{it} + \delta_{32}(UWDL)_{it} + \delta_{33}(UWP)_{it} + \alpha \\
&+ v_{it}
\end{aligned}$$

Where $i = 20$ and $t = [2004; 2022]$. The results are showed in the Table 3.

Table 3. Results of the panel data with instrumental variables.

Fixed-effects TSLS, using 337 observations, dependent variable NII			
	Coefficient	Standard Error	z
Constant	1.35905***	0.257792	5.272
ILS	-0.0257926***	0.00659064	-3.914
RBN	-0.0121140***	0.00415055	-2.919

ETF5	0.0650885***	0.0126524	5.144
OEE	0.0373954***	0.0114335	3.271
IE	0.298406***	0.0181051	16.48
EWH	-0.0884321***	0.0169215	-5.226

Statistics and Tests	SSR = 126.171
	  sigma-hat = 0.635919 (df = 312)
	  R-squared = corr(y, yhat)^2 = 0.877983
	  Included units = 19
	  Time-series length: min = 16, max = 18
	  Wald chi-square(6) = 2142.62 [0.0000]
	  Null hypothesis: The groups have a common intercept
	  Test statistic: F(18, 312) = 13.2219 [0.0000]		

However, we also tested the same results with panel data with fixed effects, panel data with random effects (Table 4).

Table 4. Panel Data with Fixed Effects and Random Effects.

	Fixed-effects, using 361 observations			Random-effects (GLS), using 361 observations		
	Coefficient	*Std. Error*	*t-ratio*	Coefficient	Std. Error	z
Constant	1.61***	0.20	7.869	1.61***	0.26	6.196
ILS	-0.02***	0.004	-5.185	-0.02***	0.004	-5.499
RBN	-0.009***	0.003	-2.737	-0.007**	0.003	-2.156
ETF5	0.05***	0.009	5.833	0.05***	0.009	5.896
OEE	0.03***	0.007	4.106	0.02***	0.007	4.005
IE	0.27***	0.01	20.84	0.2***	0.012	20.89
EWH	-0.09***	0.01	-6.914	-0.09***	0.013	-7.030
Statistics	Mean dependent var		4.570360	Mean dependent var		4.570360
	Sum squared resid		131.2132	Sum squared resid		278.9492
	LSDV R-squared		0.897893	Log-likelihood		-465.6952
	LSDV F(24, 336)		123.1107	Schwarz criterion		972.6125
	Log-likelihood		-329.5610	Rho		0.357837
	Schwarz criterion		806.3440	S.D. dependent var		1.889336
	rho		0.357837	S.E. of regression		0.886438
	S.D. dependent var		1.889336	Akaike criterion		945.3903
	S.E. of regression		0.624912	Hannan-Quinn		956.2133
	Within R-squared		0.872357	Durbin-Watson		1.212105
	P-value(F)		1.1e-150			
	Akaike criterion		709.1220			
	Hannan-Quinn		747.7753			
	Durbin-Watson		1.212105			

Tests	Joint test on named regressors -
	Test statistic: F(6, 336) = 382.724
	with p-value = P(F(6, 336) > 382.724) =
	7.02838e-147
				'Between' variance = 0.480192
				    'Within' variance = 0.363472
				        theta used for quasi-demeaning = 0.804265
				Joint test on named regressors -
				 Asymptotic test statistic: Chi-square(6) = 2324.38		

	Test for differing group intercepts - Null hypothesis: The groups have a common intercept Test statistic: $F(18, 336) = 13.205$ with p-value = $P(F(18, 336) > 13.205) =$ 1.67587e-29	with p-value = 0 Breusch-Pagan test - Null hypothesis: Variance of the unit-specific error = 0 Asymptotic test statistic: Chi-square(1) = 262.866 with p-value = 4.07192e-59 Hausman test - Null hypothesis: GLS estimates are consistent Asymptotic test statistic: Chi-square(6) = 14.132 with p-value = 0.0281954

Negative relationship between net income inequality (S80/S20) and inadequate literacy skills. The S80/S20 correlation of net income inequality with the poor skills of literacy among the regions of Italy signifies a critical socio-economic factor in that the higher the income disparities, the lower the skills of literacy proficiency. The S80/S20 income ratio of the richest 20% against the income of the 20% of the population that is the poorest signifies the intensity of the inequality. The higher the S80/S20 ratio, the higher the economic disparities that normally correlate with the disparities of good education facilities and the chance of gaining skills. In Italy, the income disparities also rank higher among the countries with lower rates of literacy, further confirming the thesis that the state of schooling has a profound effect on the state of the economy. Low literacy skills also reinforce economic exclusion and constrict the provision of good employment and further reinforce income disparities. The traditionally poorer countries of the south also rank higher in disparities and lower rates of literacy relative to the richer north of the peninsula. This vicious cycle also maintains that interventions that try to abate income disparities—such as progressive taxation, increased expenditure on schooling and specially targeted programs of literacy—may at the same time boost rates of literacy and economic equitability. The negative correlation also highlights the overarching issue of social mobility with persons with poor literacy skills not expected to move further up the economic hierarchy and reinforcing cycles of intergenerational poverty (Figure 8). The solution of the disparities will call for a two-dimensional approach based on economic redistribution and investment in education such that the acquisition of literacy becomes an antidote against inequality and not a depiction of it (Basile et al., 2022; Mussida and Parisi, 2020; Stella et al., 2020).

Figure 8. Negative relationship between net income inequality (S80/S20) and inadequate literacy skills.

The negative relationship between net income inequality (S80/S20) and reading books and newspapers.
The inverse correlation between net income inequality (S80/S20) and the culture of book and newspaper reading among the different regions of Italy emphasizes the larger economic and social cleavage that influences the participation in culture and education. The S80/S20 ratio of income of the richest 20% compared with the poorest 20% is an invaluable economic disparity index. The higher the ratio, the higher the inequality and the lower the availability of educational facilities and culture and the participation in them such as the culture of reading. In more income-differentiated countries, such as southern Italy, lower-income households may be exposed to economic and social barriers that discourage the acquisition of regular reading behavior. Books and periodicals may not be necessarily expensive but require a certain economic and cultural capital that may not be dispersed widely in lower-income economic areas. Lower-income households may spend money on the necessities of life and not necessarily on the consumption of culture and public investment in schools and library programs may be lower in such areas. This leads lower-income individuals not attaining good reading behavior and further deprives them of the chance of acquiring knowledge and the development of critical analytical skills and future employment opportunities. Conversely, lower-inequality-income areas, the north in general, have higher rates of literacy and provision of educational facilities that foster a culture of reading. Societies with higher income and lower inequality also prioritize intellectual engagement and invest in schools, libraries, and culture that foster a love of reading. This is a cycle of reinforcement: lower inequality leads to higher education and consumption of culture and a highly educated population leads to economic mobility and lower inequality in the future. Ultimately, the answer to the ill correlation of income and reading rests with targeted interventions that widen education facilities, foster literacy programs, and keep economic barriers at bay and allow participation in culture (Figure 9). Public library strengthening, the reducing of the cost of books, and the fostering of a culture of reading at an early age will aid in bridging the gulf between the economic and social disparities and the intellectual participation of the masses (Rossi, et al., 2024; Tubadji et al., 2022).

Figure 9. The negative relationship between net income inequality (S80/S20) and reading books and newspapers.

The negative relationship between net income inequality (S80/S20) and employees who work from home. The negative relationship between net income inequality (S80/S20) and the percentage of employees who work from home across Italian regions reflects the structural and socio-economic disparities that influence labor market opportunities. The S80/S20 ratio, which compares the income share of the top 20% to the bottom 20%, serves as a key indicator of inequality. A higher ratio indicates greater economic disparity, which often correlates with limited access to remote work opportunities, particularly in less affluent regions. In regions with high income inequality, such as those in southern Italy, the prevalence of remote work tends to be lower. This is due to several interrelated factors, including lower levels of digital infrastructure, a greater concentration of jobs in low-wage, manual, or service-oriented sectors that do not allow for remote work, and lower educational attainment, which limits access to high-skilled professions that offer teleworking options. Additionally, economic inequality often translates into disparities in technological access, with lower-income households less likely to have stable internet connections or adequate workspaces at home. Conversely, in regions with lower income inequality, often in the north, a more developed economic structure fosters greater opportunities for remote work. These areas tend to have stronger digital infrastructure, a higher concentration of knowledge-based industries, and a workforce with higher educational qualifications, all of which facilitate teleworking. Furthermore, companies in wealthier regions are more likely to implement flexible work arrangements, recognizing the productivity benefits of remote work and investing in the necessary digital tools (Figure 10). The negative correlation suggests that income inequality not only limits economic mobility but also affects adaptability to modern work trends. Addressing these disparities requires policies that enhance digital

access, improve education and training for remote-friendly jobs, and support businesses in less developed regions to integrate teleworking solutions, ensuring a more equitable distribution of remote work opportunities across Italy (Bonacini et al., 2020; Rossi et al., 2024; Bollani et al., 2023).

Figure 10. The negative relationship between net income inequality (S80/S20) and employees who work from home.

The positive relationship between net income inequality (S80/S20) and employed in fixed-term jobs for at least 5 years. The correlation of the net income inequality of S80/S20 with the percentage of employees with at least five years of experience in fixed-term positions among the Italian regions underscores the structural disparities of the labor market that sustain economic disparities. The S80/S20 ratio of income share of the richest 20% relative to the poorest 20% of the population is an excellent gauge of disparities in income. The higher the ratio, the higher the economic disparities and often the correlation with insecure employment forms such as long-term temporary positions. The higher the income disparities in the region and especially in southern Italy, the higher the employment insecurities of the labor market and the lower the economic development and the higher the use of temporary employment forms. Most of the workers stay in temporary positions a long time because the possibilities of permanent employment are limited and this limits the possibilities of gaining economic stability. This long-lasting precariness diminishes the possibilities of receiving benefits such as employment stability, occupational careers and stable income and increases economic disparities and reduces social mobility. Furthermore, companies in the higher-inequality regions might be inclined not to offer permanent positions because of economic uncertainty and this results in a vicious cycle of the employment situation

with temporary employment forms turning into a standard instead of a transition toward stable employment forms. The opposite happens in the lower-inequality regions and this often happens in the north of Italy with stronger economic development and a higher development of the employment system. The employees get the opportunity to move toward permanent employment forms and the occurrence of temporary employment forms lasting a long period of time diminishes. This results in a reinforcing cycle of economic disparities leading toward employment insecurities that sustain economic disparities. To tackle this problem, we need policies that foster employment stability, push employers toward providing permanent employment contracts, and lower the structural obstacles that entrap employees in temporary employment (Figure 11). Enhancing the strength of workers' rights, improving the training of workers' skills, and rewarding permanent employment contracts will help minimize the ill consequences of this correlation (Hoffmann et al., 2022; Checchi et al., 2024; Hoffmann et al., 2021).

Figure 11. The positive relationship between net income inequality (S80/S20) and employed in fixed-term jobs for at least 5 years.

The positive relationship between net income inequality (S80/S20) and irregular employees. The S80/S20 correlation with the prevalence of irregular workers among the various zones of Italy encapsulates the entrenched economic and employment disparities that drive social and economic volatility. The S80/S20 ratio that pits the income of the top 20% against that of the lower 20% of the population is a leading index of inequality. The higher the S80/S20 ratio, the higher the economic inequality and the higher the rates of informality and irregular employment that typically prevail in the lower economic zones. In the countries with higher disparities in income, such as southern Italy, the

working environment also becomes insecure with the majority of workers working informally. The employment does not get any guarantees of legality, social security coverage and stable income and instead opens the workers to exploitation and economic uncertainty. The economic necessity forces a majority of individuals and especially the lower educated individuals to take the non-regular employment and it becomes the sole source of subsistence for them. This cycle of poverty and exclusion follows with the informally working individuals earning lower income and with limited chances of advancing in careers and with higher employment uncertainty. The companies working in the countries with lower economic development and higher disparities may be willing to utilize the non-regular workers in an effort of reducing the expenditure and avoiding the regulations and further entrenching the non-regular sector of the economy. By contrast, in lower-income-in-inequality areas typically found in the north, higher economic development and stronger regulations of the workforce, coupled with a higher degree of a formal sector of employment, yield higher employment stability. Strong regulations of the workforce and higher economic development discourage non-standard employment and offer employees better working conditions and legal rights. Mitigating the prevalence of nonstandard employment will include redressing economic disparities and the susceptibility of the employment sector (Figure 12). Increased labor checks, widening the scope of standard employment prospects, and incentives that will motivate firms to transition toward regular employment deals will ease the deleterious consequences of the two variables' correlation (Hoffmann et al., 2021; Checchi et al., 2024; Carta and De Philippis, 2021).

Figure 12. The positive relationship between net income inequality (S80/S20) and irregular employees

The positive relationship between net income inequality (S80/S20) and employees who work from home. The working-at-home ratio correlation with the net income gap (S80/S20) of the Italian regions suggests that the opportunity of working at home may be within the reach of the higher-income earners and further amplify the existing income disparities. This tendency will be strongest in the developed economic regions of Lombardia, Lazio and Liguria with higher income disparities and higher take-up rates of working at home. One of the primary reasons for this correlation lies with the types of employment that offer the opportunity of working remotely. Jobs with technology, consulting, finances, and professional services—characteristics of higher-income workers—lend themselves better toward working remotely. Low-income workers in manufacturing, retailing, the hospitality sector, and the provision of individual services also offer few options of working remotely and therefore reinforce the economic cleavages. The higher the proportion of workers working remotely within an area, the greater the income disparities will be because the employment types that offer working remotely offer higher compensation, higher employment stability, and higher discretion within the workplace. Moreover, access to remote work is influenced by digital infrastructure and educational attainment. Wealthier regions tend to have better broadband connectivity and a higher percentage of highly educated workers, making teleworking more feasible. In contrast, less developed regions, particularly in the South, face digital divide issues, limiting remote work options for lower-income groups. While positives of working remotely such as a better balance between life and work and lower travel expenses do exist, uneven distribution will be likely to generate lasting social and economic disparities (Figure 13). Policymakers will be required to put emphasis on the development of the digital infrastructure and training lower-income employees in working remotely and provision inclusive of the prospects of teleworking in an attempt not to further widen income disparities among the different zones of Italy (Bonacini et al., 2020; Bollani et al., 2023; Rossi et al., 2024).

Figure 13. The positive relationship between net income inequality (S80/S20) and employees who work from home

Positive Relationship Between Net Income Inequality (S80/S20) and Employees Who Work From Home

7. A Confrontation Among Machine Learning Algorithms

We compare the performance of seven regression algorithms: Decision Tree Regression, K-Nearest Neighbors (KNN) Regression, Linear Regression, Neural Network Regression, Random Forest Regression, Regularized Linear Regression, and Support Vector Machine (SVM) Regression, across five key evaluation metrics: Mean Squared Error (MSE), Scaled MSE, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE/MAD), and R^2 (coefficient of determination). After analyzing the results, it becomes evident that Random Forest Regression outperforms all other models in terms of accuracy and error reduction (Table 5).

Table 5. Performance of machine learning algorithms.

Machine Learning Algorithms	MSE	MSE (Scaled)	RMSE	MAE/MAD	R^2
Decision Tree	0.461	0.111	0.679	0.465	0.89
K-Nearest Neighbors	0.629	0.167	0.793	0.527	0.838
Linear Regression	0.641	0.189	0.801	0.623	0.836
Neural Network	0.734	0.113	0.857	0.559	0.888
Random Forest	0.273	0.068	0.522	0.407	0.932
Regularized Linear	0.529	0.107	0.727	0.575	0.894

Support Vector Machine	0.746	0.182	0.864	0.691	0.824

To understand the effectiveness of each regression model, we analyze the following key performance metrics. MSE measures the average squared difference between actual and predicted values, where lower values indicate better model performance. Scaled MSE is a normalized version of MSE that allows better comparison across models. RMSE is the square root of MSE, providing an interpretable error measurement in the same unit as the target variable. MAE measures the average absolute difference between predictions and actual values. R^2 represents how well the model explains the variance in the data, with higher values indicating better fit.

Decision Tree Regression demonstrates a moderate level of performance with an MSE of 0.461, RMSE of 0.679, and MAE of 0.465. While its R^2 value of 0.89 is relatively high, it is still outperformed by other models in terms of accuracy and stability. Decision Trees tend to overfit when data complexity increases, which can limit their generalization ability. The KNN Regression model shows a slightly higher error rate, with MSE of 0.629, RMSE of 0.793, and MAE of 0.527. Its R^2 value is 0.838, indicating a weaker ability to explain variance compared to Decision Trees and other complex models. KNN regression is highly dependent on the choice of neighbors and can struggle with high-dimensional data. Linear Regression, being a simple model, results in MSE of 0.641, RMSE of 0.801, and MAE of 0.623. Its R^2 score of 0.836 suggests that it captures some patterns but fails to model complex relationships. This model assumes linear relationships, which may not hold true for many real-world datasets. The Neural Network model offers a slightly improved performance over KNN and Linear Regression, with an MSE of 0.734, RMSE of 0.857, and MAE of 0.559. The R^2 value of 0.888 indicates good variance explanation, but its error rates remain higher than other models, suggesting potential overfitting or inefficient parameter tuning. Regularized Linear Regression performs better than standard Linear Regression, reducing overfitting by penalizing large coefficients. It achieves an MSE of 0.529, RMSE of 0.727, and MAE of 0.575. The R^2 score of 0.894 suggests improved predictive capability. However, it is still outperformed by ensemble-based methods like Random Forest. SVM Regression exhibits the highest error rates among all models, with an MSE of 0.746, RMSE of 0.864, and MAE of 0.691. The R^2 score of 0.824 is the lowest in the dataset, indicating that the model struggles to explain the variance in the data. This could be due to poor kernel selection or the inherent limitations of SVM regression in capturing highly nonlinear relationships (Subbotin, 2020; Dawat, 2023).

Among all models, Random Forest Regression demonstrates superior performance with the lowest errors and the highest variance explanation. It has the lowest MSE of 0.273, significantly lower than all other models, indicating minimal squared error. The lowest Scaled MSE of 0.068 confirms its robustness in reducing prediction errors. The lowest RMSE of 0.522 suggests that it has the least deviation from actual values. The lowest MAE of 0.407 indicates that on average, its predictions are closest to the true values. The highest R^2 of 0.932 means it explains 93.2% of the variance in the dataset, significantly outperforming other models. The Random Forest Regression model excels because it leverages an ensemble of decision trees, reducing overfitting while capturing complex patterns in the data. It works by training multiple decision trees on different random subsets of the data, averaging predictions to reduce variance and improve generalization, and handling both linear and nonlinear relationships effectively.

In real-world applications, Random Forest Regression is an ideal choice when high accuracy is required and errors must be minimized, when there is a mix of linear and nonlinear relationships in the dataset, when the dataset is large as Random Forest can efficiently handle large volumes of data, and when feature importance needs to be evaluated as it provides insights into which variables are most influential in

predictions. Despite these advantages, Random Forest Regression can be computationally expensive, especially with very large datasets. However, given its superior predictive performance in this scenario, it remains the best choice. After analyzing the performance of multiple regression models, Random Forest Regression clearly emerges as the best-performing model due to its lowest error rates across all metrics, highest explanatory power with R^2 of 0.932, and ability to handle complex data structures and relationships. While simpler models like Linear Regression and Decision Trees provide interpretability, they do not achieve the same level of accuracy as Random Forest. Likewise, Neural Networks and SVM Regression fail to outperform ensemble methods in this case. Thus, for optimal predictive accuracy and generalization, Random Forest Regression is the best model to use in this dataset (Maia et al., 2021; Thohari and Ramadhani, 2022; Gooljar et al., 2023).

The application of the Random Forest Regression algorithm in the prediction of net income inequality (S80/S20) among the Italian regions returns valuable results regarding the effect of multiple variables of the labor situation and education on income inequality. The Feature Importance Measures will identify the relative significance of the predictors in affecting economic inequality and the complex interaction among the factors of the employment situation, education, and economic distribution. Through the evaluation of the Mean Decrease in Accuracy, Total Increase in Node Purity, and Mean Dropout Loss, we will be able to estimate the relative significance of each of the features and the effect it has on net income inequality (Table 6).

Table 6. Feature importance metrics.

	Mean decrease in accuracy	Total increase in node purity	Mean dropout loss
Irregular employees	3.135	164.019	1.467
Overeducated employees	0.952	69.090	0.814
Employees who work from home	0.864	55.188	0.748
Inadequate literacy skills (students in third grade of lower secondary school)	0.725	46.010	0.653
Reading books and newspapers	0.162	30.783	0.649
Employed in fixed-term jobs for at least 5 years	0.440	15.849	0.560
Note. Mean dropout loss (defined as root mean squared error (RMSE)) is based on 50 permutations.			

Irregular employees stand out as the most relevant factor in income inequality with the highest Mean Decrease in Accuracy of 3.135, Total Increase in Node Purity of 164.019, and Mean Dropout Loss of 1.467. The findings reveal that informal employment has a major influence in widening income disparities. The predominance of irregular employment in Italy and specifically in the southern part of the country has been a long-standing factor associated with employment insecurity, lower earning power, and limited access to social protections. Those in the informal employment sector earn lower income and limited prospects of advancing professionally and therefore widening the income disparities between the richest and the poorest. The highest values of the entire set of the three feature importance metrics imply that any intervention of a policy in reducing income disparities needs to be focused specifically at addressing the issue of the informal employment sector with stronger employment security and stronger enforcement of the labor legislation.

The second primary driver of this factor includes employees with a Mean Decrease in Accuracy of 0.952, Total Increase in Node Purity of 69.090, and Mean Dropout Loss of 0.814. This indicates that the mismatch of employees' educational attainment and the positions they take has a tremendous influence on income inequality. Overeducation has a tendency of triggering underemployment, lower income, and stagnation of careers that impede the economic mobility of individuals. This problem mainly affects the youths graduating in Italy and gaining employment that matches the skills and qualifications they possess, and this leads to a situation that does not necessarily equate higher education with higher income. This problem raises economic inequality with highly educated individuals working in low-income careers and individuals with unfavorable positions having limited options. This problem requires the matching of education with the needs of the workforce such that the graduating students at the university level will be able to acquire employment that matches the skills and qualifications they possess.

One of the primary factors includes home-working employees with a Mean Reduction in Accuracy of 0.864, Total Rise in Node Purity of 55.188, and Mean Dropout Loss of 0.748. The COVID-19 outbreak has changed income distribution patterns with the advent of home working. Higher income and higher-flexibility employment possibilities of home working and skilled sector employment notwithstanding, home working increases economic disparities within and among regions with residents of rural and depressed economic areas not necessarily having the option of home working. The rankings of the top home working features imply that home working has a positive correlation with income inequality and that this may be the reason that higher-income earners get the greatest benefit of home working possibilities and lower-income earners, particularly with manual and service-type employment, get excluded from this opportunity. This observation underscores the need for the development of policies that augment digital inclusion and equitable provision of home working possibilities and include the development of broadband facilities in underserved areas and the provision of training in digital skills. Another vital educational factor includes the students at the third grade of lower secondary with a Mean Decrease in Accuracy of 0.725, Total Increase in Node Purity of 46.010, and Mean Dropout Loss of 0.653. This reflects that poor skills at an early age play a significant role in economic disparities in the long term. The proficiency of skills at literacy plays a major influencing factor in future employment opportunity, income levels, and economic mobility in general. The areas with students that fare badly at the primary skills of literacy will be the areas that will bear the higher income disparities because the individuals with lower education will be at a disadvantage of finding stable and good-remunerated employment. The correlation of the strong linkage of income disparities and literacy emphasizes the necessity of the improvement of the programs of the early education, particularly in the economically depressed areas, so that the students will be able to construct the foundation skills necessary for higher education and employment opportunity (Wen et al., 2024; Sieck et al., 2021; Xu et al., 2024; Wang et al., 2024).

Reading and book and newspaper consumption play a relatively lower role of significance in the prediction of income inequality with a Mean Reduction in Accuracy of 0.162, Total Node Purity Increase of 30.783, and Mean Dropout Loss of 0.649. Yet this factor does influence economic disparities. The consumption of reading is typically associated with higher education and critical reasoning skills that further influence employment and income rates. Those who partake in regular consumption of reading will get better cognitive abilities, be better educated about economic and political patterns, and be at a higher chance of career development. The lower relative scores of the features imply that although the consumption of reading does influence income inequality, it does so in a less direct manner compared to employment market conditions and the level of education. Yet the instillation of the consumption of reading at an early age will still yield long-lasting consequences of reducing educational disparities and making the workforce competitive.

The last factor, used in temporary employment positions lasting at least five years, has a Mean Reduction in Accuracy of 0.440, Total Gain in Node Purity of 15.849, and Mean Dropout Loss of 0.560. Even if this factor has lower relative importance compared with the rest of the factors, it also plays a part in income inequality through the emphasis it places on employment precarité and employment market volatility. Fixed-term employment agreements, widely adopted in Italy, constrain the development of careers, diminish the stability of employment positions, and induce wage stagnation. Those workers who stay in temporary employment positions for a long duration encounter obstacles in the negotiation of higher compensation, the receipt of fringe benefits, and the attainment of stable economic future prospects. The finding of this study indicates that income-reducing inequality policies should also be addressed toward the enhancement of employment stability and workers' rights in temporary employment agreements. In conclusion, the Random Forest Regression results demonstrate that labor market conditions and educational disparities play a fundamental role in shaping income inequality across Italian regions. The most influential factor is irregular employment, which significantly widens the income gap by limiting access to stable wages and social protections. Other key contributors include overeducation, remote work, and early literacy skills, all of which reflect structural issues within the labor market and education system. While reading habits and fixed-term employment have relatively lower importance scores, they still contribute to long-term income disparities. These findings emphasize the need for targeted policy interventions, including strengthening labor protections, improving educational access, and promoting economic inclusion to reduce inequality and foster a more balanced income distribution. Addressing these structural challenges will be essential in creating a more equitable economic environment where all individuals, regardless of their background, have equal opportunities for financial success and social mobility (Fauser and Gebel, 2023).

8. Policy Implications

Reducing income inequality across Italian regions requires a comprehensive policy strategy that integrates education, labor market reforms, infrastructure investments, and digital transformation. Given the persistent North-South economic divide, targeted interventions should aim to narrow regional disparities and promote sustainable economic convergence. This study highlights the critical role of labor market conditions, education, and digitalization in shaping income inequality, emphasizing the need for policies that address these key factors while fostering inclusive economic growth. One of the major issues of the South of the nation and the overall Italian labor market includes the pervasive use of the non-regular employment sector. The scope of non-regular employment has a vital effect on income inequality because the workers of the non-regular sector face employment uncertainty, the absence of social guarantees, and restricted social ascendance prospects. Policymakers should enhance the rules of the labor market and the enforcement of the prohibition of non-regular employment and incentivize companies that regularize workers. Tax and social contribution reforms may convince companies to hire registered workers and enhance the economic stability of workers and the business sector in general. Fixed-term employment is also a primary source of income inequality, especially in areas with a preponderance of temporary employment in the workforce. Sustained use of short-term employment stunts wage increases, hinders occupational development, and increases economic uncertainty. Policy should be used against this effect by incentivizing the transition of temporary employment into permanent employment. This might be done with tax incentives to companies that issue long-term employment contracts or with tighter controls on the rollover of temporary employment contracts. Active labor market programs, including vocation training and retraining programs, also need to be expanded in

order to prepare workers with skills that will allow them to be hired in stable, high-value added employment (Berton et al., 2023; Davidescu et al., 2024).

Educational attainment has a major influence in the determination of income distribution. The highly negative correlation of income inequality with skills in literacy emphasizes the necessity of large investment in education, especially in the developing world. The placement of a premium at the primary level with a provision of good foundation skills in numeracy and literacy and the provision of focused interventions at the secondary and higher education will fill the skills shortages that limit economic mobility. The provision of additional scholarships programs and the construction of good schools and the training of educators in the poor neighbourhood will help narrow the education gap and maximize income distribution in the long term. Regional income disparities also arise because of the digital divide. The South has lower coverage of broadband internet, lower digital proficiency rates, and lower technology-intensive firm densities relative to the North. This has to be bridged if inclusive economic development has to be fostered. Broadband development will require additional investment in rural and disadvantaged economic areas and has to be a government top-priority area of investment. Public-private sector initiatives will be critical in financing and expediting the installation of the necessitated digital infrastructure. Digital reskilling initiatives also need to be established so that the workforce of the lagging areas has the competences required to be part of the digital economy. Remote working options that have been gradually gaining traction in the aftermath of the pandemic also offer a solution toward bridging the gaps at the regional scale. The scope of working remotely has a strong correlation with income and educational attainment. Policymakers should prioritize expanding the scope of initiatives toward digital inclusion, particularly in the South. The push toward companies having remote working centers in the lower development zones will help stimulate the respective economies and generate employment that does not get restricted based on geographical limitations (Figure 14). Financial incentives granted to companies that hire remote workers based in lower development zones will also help integrate the economy further (Hollman et al., 2021; Byrum, 2023; Wang et al., 2024; Zhang et al., 2024).

Figure 14. How to reduce income inequality across Italian regions.
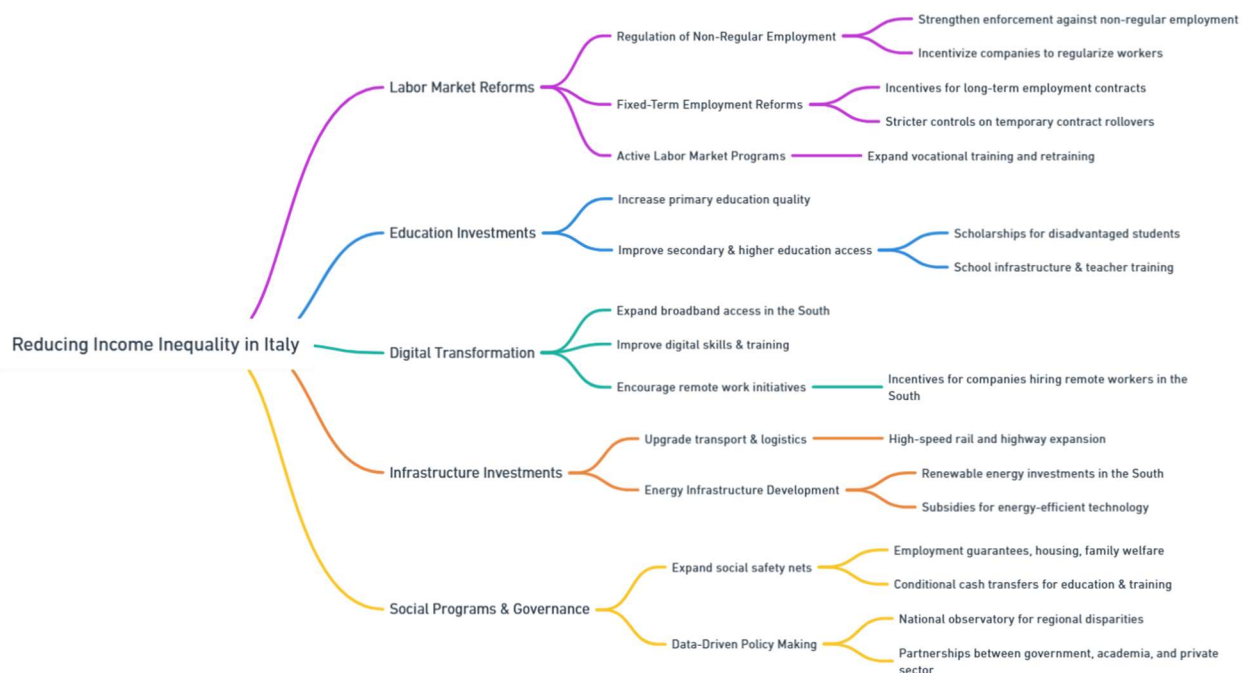
Regional economic development is a major driver of investment in infrastructure. The Japanese and Chinese experience has proven that large public investment in transport, energy, and industrial infrastructure reduces income disparities within the regions substantially. The economic development of Italy's South has been impeded by persisting infrastructure deficiencies. Poor transport connections, public services deficits, and outdated energy facilities hinder economic development. The solution lies in a unified investment strategy that upgrades transport systems, improves the efficiency of energy and public services in the disadvantaged areas. Higher investment in the construction of a high-speed rail network and highways and the development of logistics centers has the potential of further economic integration of the South with the North and bringing business in the disadvantaged areas closer to larger markets. Energy policy also has a vital influence on economic disparities. The transition to renewable energy has the chance of spurring economic development in southern Italy, with abundant solar and wind energy resources. Nevertheless, investment in the development of renewable energy facilities has been disproportionately skewed toward the North. Policy incentives need to be put in place that will stimulate the construction of clean energy projects in the South and generate employment and lower the energy bills of companies and households. Support and incentives in the forms of subsidies and grants also need to be granted toward the use of energy-efficient technologies in the economically disadvantaged zones (Komornicki and Goliszek, 2023; Rosik and Wójcik, 2022; Marino et al., 2023).

Social programs should be supported with economic interventions so that the poor sections of the population receive the necessary support. Expansion of social safety nets such as employment guarantees, housing subsidies and family welfare programs will ease the short-run effect of income disparities. They should be designed in a way that stimulates economic inclusion and not dependency in the long term. Conditional cash transfers that provide cash transfers against education and training enrollment may be an effective intervention in reducing poverty and fostering human capital development. Finally, reducing regional income disparities needs a long-term effort at policy coordination and evidence-informed decision-making. Policymakers need to take a data-informed approach with constant observation of regional economic patterns and evaluation of the effect of interventions. The development of a national observatory of regional disparities might help in the gathering and analysis of economic statistics that will allow a more accurate and flexible policy intervention. Furthermore, the improvement of the partnership among government institutions, academe and private sector stakeholders will help in making the intervention of the government more effective and development efforts focused and relevant to economic needs. In conclusion, reducing regional income inequality in Italy necessitates a comprehensive and integrated policy framework that addresses labor market dynamics, educational disparities, digital transformation, infrastructure development, and governance reforms. Targeted investments in education, job creation, and digital inclusion are essential for fostering sustainable economic convergence between the North and the South. By adopting a strategic and evidence-based approach, policymakers can create an economic environment that promotes equitable growth, enhances social mobility, and ensures that all regions of Italy have the opportunity to thrive (Hyee et al., 2020; Mazzeo Rinaldi and Leone, 2023).

9. Conclusions

This analysis provides a comprehensive overview of income disparities among the Italian territories with the help of instrumental variable panel data models, k-means clustering and machine learning models. The findings reaffirm the long-standing North-South cleavage of Italy with the greatest disparities

observed in the least developed territories of Campania, Calabria, and Sicily. Through the use of econometric techniques, we identified the primary factors of income disparities such as the informality of employment, temporary employment, overeducation and the digital divide. The underlying factors persistently hindering economic convergence and entrenching the disparities among the territories call for the need of sector-targeting interventions addressing the inefficiencies of the labor market and the improvement of digital and educational opportunities.

Cluster analysis confirms that economic territorial segregation of the Italian territories has the north and centre representing a high-income convergence cluster and southern territories trapped in a cycle of low-income equilibrium. The tendency has the world patterns of the experience of such countries with territorial disparities such as Indonesia with income distribution highly dependent on the development of the infrastructure and the state of the employment market. Italy's experience has the additional dimension of longstanding institutional and historical disparities that began with the period of the aftermath of the unification and further consolidated the economic cleavages.

Machine learning techniques, that is the application of a random regression forest, were applied in an effort to estimate income inequality patterns and rank the top variables. Outcomes reveal non-registered employment as the greatest driver of territorial disparities closely followed by overeducation and the absence of digital infrastructure. The correlation of low penetration with top income inequality emphasizes the need to adopt digital inclusion in an effort to mitigate economic disparities. Broadband availability and the development of digital skills training programs, especially in the Mezzogiorno, will dramatically enhance employment and economic opportunity prospects.

Income disparities are conditioned a great extent by the forces of the labor market. The use of temporary employment and the predominance of informality within the southern countries constrain employment stability and wage increases and widen economic disparities. The results place a premium on reforms of the labor market that will enhance stable employment prospects. Those that advance permanent employment agreements, reduce the use of insecure employment forms and enhance the provision of training programs of a vocational character will be instrumental in inducing economic stability and the elimination of income disparities. Education becomes yet another factor that affects income disparities. The study discovers a clear-cut correlation between the illiteracy rates and income disparities and emphasizes the need for heavy investment in schooling. The countries with higher illiteracy rates of good schooling exhibit lower illiteracy rates of disparities and the countries with a weak schooling system exhibit entrenched economic disparities. Widening the scope of good schooling and the provision of higher schooling and making the curricula of the latter suit the needs of the employment sector will be instrumental in filling the skills shortage and inducing social mobility.

Development of the infrastructure has been the major driver of economic convergence. The Japanese experience and that of China demonstrate the power of investment in transport and energy and telecommunications infrastructure in bridging the disparities at the regional level. The experience of Italy varies with the entrenched South-to-North infrastructure gap acting as a constraint in the economic development of the South and the development of the manufacturing sector and the availability of linkage with larger markets acting as a constraint. The solution will be an integrated investment strategy that pursues the modernization of transport facilities and the development of programs of renewable energy and the improvement of telecommunications connections. This will be crucial in the development of the regions and the facilitation of economic integration in the long term.

Further, this study highlights the importance of the synergy of the econometric and machine learning techniques of income disparities analysis at the state level. The application of state of the art quantitative techniques will allow a deep and empirical understanding of the underlying factors of income disparities. The application of the use of predictive models will specifically provide policymakers with a useful asset

with which they will be able to design interventions that target the driver of disparities and construct economic development that will be durable. The solution of Italy's income disparities at the regional level will also call for a comprehensive and empirically based framework of policy. Investments in the sector of education, the sector of the digital economy, and the sector of the workforce will be instrumental in fostering economic convergence and alleviating disparities. The improvement of institutions' efficiency and the reinforcement of the government mechanisms will also play an instrumental role in bridging the South-North economic gap. Through the application of evidence-informed and strategic policies, Italy will be in a position to build a balanced and inclusive economic environment that will allow the entire country to take advantage of equitable development and growth opportunities.

10. References

Acheampong, A. O., Dzator, J., & Shahbaz, M. (2021). Empowering the powerless: does access to energy improve income inequality?. Energy Economics, 99, 105288.

Acheampong, A. O., Shahbaz, M., Dzator, J., & Jiao, Z. (2022). Effects of income inequality and governance on energy poverty alleviation: Implications for sustainable development policy. Utilities Policy, 78, 101403.

Aginta, H., Gunawan, A. B., & Mendez, C. (2023). Regional income disparities and convergence clubs in Indonesia: new district-level evidence. Journal of the Asia Pacific Economy, 28(1), 101-132.

Akita, T., & Alisjahbana, A. S. (2023). The Initial Impacts of the COVID-19 Pandemic on Regional Economies in Indonesia: Structural Changes and Regional Income Inequality. Sustainability, 15(18), 13709.

Aresu, F., Marrocu, E., & Paci, R. (2023). Public capital and institutions' quality in the Italian regions. Journal of Regional Science, 63(5), 1284-1308.

Asongu, S. A., & Odhiambo, N. M. (2021). Inequality, finance and renewable energy consumption in Sub-Saharan Africa. Renewable Energy, 165, 678-688.

Asso, P. F. (2023). New perspectives on old inequalities: Italy's north–south divide. In Inequalities, territorial politics, Nationalism (pp. 22-40). Routledge.

Baek, I., Noh, S., & Ahn, J. (2024). Does income inequality move together across the world?. Applied Economics Letters, 31(13), 1195-1200.

Basile, R., Ciccarelli, C., & Groote, P. (2022). The legacy of literacy: evidence from Italian regions. Regional Studies, 56(5), 794-807.

Berton, F., Pacelli, L., Quaranta, R., & Trentini, F. (2023). Patterns of labor market reforms: a regional approach to the Italian 'Jobs Act'. Sinappsi, 13, 50-67.

Bollani, L., Di Zio, S., & Fabbris, L. (2023). Chapter Remote working in Italy: Just a pandemic accident or a lesson for the future?. In ASA 2022 Data-Driven Decision Making. Firenze University Press, Genova University Press.

Bonacini, L., Gallo, G., & Scicchitano, S. (2020). All that glitters is not gold. Influence of working from home on income inequality at the time of Covid-19.

Bruns-Smith, D., Feller, A., & Nakamura, E. (2023, June). Using Supervised Learning to Estimate Inequality in the Size and Persistence of Income Shocks. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 1747-1756).

Byrum, G. (2023). Opening the Broadband Access Paradox. JCMS: Journal of Cinema and Media Studies, 62(4), 193-199.

Caragliu, A., & Del Bo, C. F. (2022). Smart cities and urban inequality. Regional Studies, 56(7), 1097-1112.

Carta, F., & De Philippis, M. (2021). The impact of the COVID-19 shock on labour income inequality: Evidence from Italy. Bank of Italy Occasional Paper, (606).

Castelló-Climent, A., & Doménech, R. (2021). Human capital and income inequality revisited. Education Economics, 29(2), 194-212.

Checchi, D., Jappelli, T., Marino, I., & Scognamiglio, A. (2024). Inequality trends in a slow-growing economy: Italy, 1990–2020. Fiscal Studies, 45(3), 377-392.

Chen, T., Gozgor, G., & Koo, C. K. (2021). Pandemics and income inequality: what do the data tell for the globalization era?. Frontiers in Public Health, 9, 674729.

Chi, Z., Lun, H., Ma, J., & Zhou, Y. (2024). Income inequality and healthcare utilization of the older adults-based on a study in three provinces and six cities in China. Frontiers in Public Health, 12, 1435162.

Condino, F. (2023). Share density-based clustering of income data. Statistical Analysis and Data Mining: The ASA Data Science Journal, 16(4), 336-347.

Culotta, F. (2021). Life expectancy heterogeneity and pension fairness: An Italian north-south divide. Risks, 9(3), 57.

Daniele, V. (2021). Territorial disparities in labour productivity, wages and prices in Italy: What does the data show?. European Urban and Regional Studies, 28(4), 431-449.

Davidescu, A. A., Lobonț, O. R., & Nae, T. M. (2024). The Fabric of Transition: Unraveling the Weave of Labor Dynamics, Economic Structures, and Innovation on Income Disparities in Central and Eastern Europe Nations. Economies, 12(3), 68.

Dawat, E. R. R. (2023). Predictive Modeling of PowerSchool Usage: Comparative Analysis of Linear Regression and Data Mining Techniques Using Student Attributes. International Journal of Research and Innovation in Social Science, 7(11), 75-85.

Dong, K., Dou, Y., & Jiang, Q. (2022). Income inequality, energy poverty, and energy efficiency: who cause who and how?. Technological Forecasting and Social Change, 179, 121622.

Fauser, S., & Gebel, M. (2023). Labor market dualism and the heterogeneous wage gap for temporary employment: a multilevel study across 30 countries. Socio-Economic Review, 21(4), 2069-2091.

Feldman, M., Guy, F., & Iammarino, S. (2021). Regional income disparities, monopoly and finance. Cambridge Journal of Regions, Economy and Society, 14(1), 25-49.

Ferto, I., & Bojnec, S. (2023). Subsidies and the income inequality in the Hungarian wine sector. Wine Economics and Policy, 12(2), 3-14.

Futagami, R. (2022). Regional Income Inequality and Allocation of Public Investment: The Japanese Experience, 1958-1986. Adjustments of economics and enterprises in a changing world, 56, 101.

Gao, J., Liu, Y., Chen, J., & Cai, Y. (2022). Demystifying the geography of income inequality in rural China: A transitional framework. Journal of Rural Studies, 93, 398-407.

Gaubert, C., Kline, P., Vergara, D., & Yagan, D. (2021, May). Trends in US spatial inequality: concentrating affluence and a democratization of poverty. In AEA Papers and proceedings (Vol. 111, pp. 520-525). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

Ghazouani, T., & Beldi, L. (2022). The impact of income inequality on carbon emissions in Asian countries: Non-parametric panel data analysis. Environmental Modeling & Assessment, 27(3), 441-459.

Giannoni, P., Palumbo, M., Pandolfini, V., & Torrigiani, C. (2024). Territorial Disparities in the Governance of Policies Promoting the School-to-Work Transition: An Analysis of the Italian Case. Education Sciences, 14(3), 260.

Ginsburgh, V., Magerman, G., & Natali, I. (2021). COVID-19 and the role of inequality in French regional departments. The European Journal of Health Economics, 22, 311-327.

Gooljar, S., Manohar, K., & Hosein, P. (2023). Performance evaluation and comparison of a new regression algorithm. arXiv preprint arXiv:2306.09105.

Guzzardi, D., Palagi, E., Roventini, A., & Santoro, A. (2024). Reconstructing income inequality in italy: New evidence and tax system implications from distributional national accounts. Journal of the European Economic Association, 22(5), 2180-2224.

Hoffmann, E. B., Malacrino, D., & Pistaferri, L. (2021). Labor market reforms and earnings dynamics: the Italian case. International Monetary Fund.

Hollman, A. K., Obermier, T. R., & Burger, P. R. (2021). Rural measures: A quantitative study of the rural digital divide. Journal of Information Policy, 11, 176-201.

Huang, X. (2024). Predictive models: regression, decision trees, and clustering. Applied and Computational Engineering, 79, 124-133.

Hyee, R., Immervoll, H., Lee, J., & Fernandez, R. (2020). How reliable are social safety nets? Value and accessibility in situations of acute economic need.

Hyun-Chool, L. E. E., & Repkine, A. (2023). Ideological Polarization and Income Inequality in the Korean Regions. Korea Journal, 63(1), 118-149.

Kashwan, K. R., & Velu, C. M. (2013). Customer segmentation using clustering and data mining techniques. International Journal of Computer Theory and Engineering, 5(6), 856.

Khan, S., & Yahong, W. (2022). Income inequality, ecological footprint, and carbon dioxide emissions in Asian developing economies: what effects what and how?. Environmental Science and Pollution Research, 29(17), 24660-24671.

Khan, S., Yahong, W., & Zeeshan, A. (2022). Impact of poverty and income inequality on the ecological footprint in Asian developing economies: Assessment of Sustainable Development Goals. Energy Reports, 8, 670-679.

Kim, Y., Sommet, N., Na, J., & Spini, D. (2022). Social class—not income inequality—predicts social and institutional trust. Social Psychological and Personality Science, 13(1), 186-198.

Kocurová, T., & Hampel, D. (2020). Inequality in the Income of the Population as a Determinant of the Country's Economic Growth. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis.

Komornicki, T., & Goliszek, S. (2023). New transport infrastructure and regional development of Central and Eastern Europe. Sustainability, 15(6), 5263.

Korotaj, T., Kurnoga, N., & Šimurina, N. (2023). Multivariate analysis of post-transition OECD countries in the context of inequality measures. Croatian operational research review, 14(1), 53-64.

Leogrande, A., Costantiello, A., & Leogrande, D. (2023). The Socio-Economic Determinants of the Number of Physicians in Italian Regions.

Li, G., Zhang, R., Feng, S., & Wang, Y. (2022). Digital finance and sustainable development: Evidence from environmental inequality in China. Business Strategy and the Environment, 31(7), 3574-3594.

Lipps, J., & Schraff, D. (2021). Regional inequality and institutional trust in Europe. European Journal of Political Research, 60(4), 892-913.

Ma, H. Can Innovation Input Help Reduce the Income Inequality in China?—An Analysis Based on Panel Data of 21 Provincial Regions.

Maia, M., Azevedo, A. R., & Ara, A. (2021). Predictive comparison between random machines and random forests. Journal of Data Science, 19(4), 593-614.

Marino, A., Pariso, P., & Picariello, M. (2023). Exploring the Economic Recovery of Italy's Regions Post-COVID-19: A focus on Energy, Services, ICT Opportunities, and the Digital Divide. International Journal of Energy Economics and Policy, 13(5), 271-280.

Mazzeo Rinaldi, F., & Leone, L. (2023). Conditional cash transfers in OECD countries: a realist synthesis. Frontiers in Sociology, 8, 1202430.

Mdingi, K., & Ho, S. Y. (2021). Literature review on income inequality and economic growth. MethodsX, 8, 101402.

Mercado, R. V., Park, C. Y., & Zhuang, J. (2023). Trends and drivers of income inequality in the Philippines, Thailand, and Viet Nam: A decomposition analysis (No. 692). ADB Economics Working Paper Series.

Munandar, T. A. (2023). K-Means Cluster Algorithm for Grouping Inequality in Regional Development.

Mussida, C., & Parisi, M. L. (2020). Features of personal income inequality before and during the crisis: An analysis of Italian regions. Regional studies.

Nassif Pires, L., Carvalho, L. B. D., & Lederman Rawet, E. (2021). Multi-dimensional inequality and COVID-19 in Brazil. Investigación económica, 80(315), 33-58.

Palagi, E., Coronese, M., Lamperti, F., & Roventini, A. (2022). Climate change and the nonlinear impact of precipitation anomalies on income inequality. Proceedings of the National Academy of Sciences, 119(43), e2203595119.

Panzera, D., & Postiglione, P. (2022). The impact of regional inequality on economic growth: a spatial econometric approach. Regional Studies, 56(5), 687-702.

Rosik, P., & Wójcik, J. (2022). Transport infrastructure and regional development: A survey of literature on wider economic and spatial impacts. Sustainability, 15(1), 548.

Rossi, R., Di Lorenzo, G., Jannini, T. B., Ossola, P., Belvederi Murri, M., Siracusano, A., & Rossi, A. (2024). The role of income inequality as an ecological determinant of mental health: A nation-wide multilevel analysis on an Italian sample. International Journal of Social Psychiatry, 70(5), 999-1003.

Saeed, S., & Siraj, T. (2024). Global Renewable Energy Infrastructure:: Pathways to Carbon Neutrality and Sustainability. Solar Energy and Sustainable Development Journal, 13(2), 183-203.

Santamato, V., Tricase, C., Faccilongo, N., Iacoviello, M., Pange, J., & Marengo, A. (2024). Machine learning for evaluating hospital mobility: an Italian case study. Applied Sciences, 14(14), 6016.

Sbardella, A., Andrea, Z., Luciano, P., & Scaramozzino, P. (2021). Behind the Italian regional divide: An economic fitness and complexity perspective. SINAPPSI, 11(2), 50-73.

Shudanko, P. (2024). How to Maintain Business Sustainability and Performance in Dynamic Global Market. Journal of Current Research in Business and Economics, 3(1), 1174-1220.

Sieck, C. J., Sheon, A., Ancker, J. S., Castek, J., Callahan, B., & Siefer, A. (2021). Digital inclusion as a social determinant of health. NPJ digital medicine, 4(1), 52.

Sotomayor, O. J. (2021). Can the minimum wage reduce poverty and inequality in the developing world? Evidence from Brazil. World Development, 138, 105182.

Stella, G. P., Filotto, U., & Cervellati, E. M. (2020). Could Financial Literacy Become a Key Variable to Examine Social and Economic Inequalities? A Study on Italian Regions. International Journal of Trade, Economics and Finance, 11(1).

Subbotin, S. (2020). Radial-basis function neural network synthesis on the basis of decision tree. Optical Memory and Neural Networks, 29(1), 7-18.

Sugiharti, L., Purwono, R., Esquivias, M. A., & Rohmawati, H. (2023). The nexus between crime rates, poverty, and income inequality: A case study of Indonesia. Economies, 11(2), 62.

Sun, Y. Y., Li, M., Lenzen, M., Malik, A., & Pomponi, F. (2022). Tourism, job vulnerability and income inequality during the COVID-19 pandemic: A global perspective. Annals of Tourism Research Empirical Insights, 3(1), 100046.

Sung, J., Qiu, Q., & Marton, J. (2021). Income Inequality and Health: New Methodology and an Application. Economics Bulletin, 41(4), 2676-2689.

Suratman, E., & Mayudi, G. (2022). IMPACT OF INFLATION AND EXCHANGE RATE ON ASEAN INCOME INEQUALITY. International Journal of Business & Society, 23(1).

Tan, Y., & Uprasen, U. (2021). Carbon neutrality potential of the ASEAN-5 countries: Implications from asymmetric effects of income inequality on renewable energy consumption. Journal of Environmental Management, 299, 113635.

Thohari, A. N. A., & Ramadhani, R. D. (2022). Performance Comparison Supervised Machine Learning Models to Predict Customer Transaction Through Social Media Ads. J. Comput. Networks, Archit. High Perform. Comput, 4(2), 116.

Tubadji, A., Gheasi, M., Crociata, A., & Odoardi, I. (2022). Cultural capital and income inequality across Italian regions. Regional Studies, 56(3), 459-475.

Ullah, A., Kui, Z., Ullah, S., Pinglu, C., & Khan, S. (2021). Sustainable utilization of financial and institutional resources in reducing income inequality and poverty. Sustainability, 13(3), 1038.

Van Ham, M., Tammaru, T., Ubarevičienė, R., & Janssen, H. (2021). Urban socio-economic segregation and income inequality: A global perspective (p. 523). Springer Nature.

Wan, G., Wang, C., Wang, J., & Zhang, X. (2022). The income inequality-CO2 emissions nexus: Transmission mechanisms. Ecological Economics, 195, 107360.

Wang, P., Li, Z., Wang, Y., & Wang, F. (2024). Unveiling the Dynamics of Educational Equity: Exploring the Third Type of Digital Divide for Primary and Secondary Schools in China. Sustainability, 16(11), 4868.

Wen, C., Xiao, Y., & Hu, B. (2024). Digital financial inclusion, industrial structure and urban–Rural income disparity: Evidence from Zhejiang Province, China. Plos one, 19(6), e0303666.

Xu, C., Han, M., Dossou, T. A. M., & Bekun, F. V. (2021). Trade openness, FDI, and income inequality: Evidence from sub-Saharan Africa. African Development Review, 33(1), 193-203.

Xu, G., Feng, L., Wang, W., & Liang, Q. (2024). Digital Financial Literacy and Rural Income Inequality. SAGE Open, 14(3), 21582440241275642.

Xu, Q., & Zhong, M. (2023). The impact of income inequity on energy consumption: The moderating role of digitalization. Journal of Environmental Management, 325, 116464.

Yang, F., Zhang, S., & Sun, C. (2020). Energy infrastructure investment and regional inequality: Evidence from China's power grid. Science of the Total Environment, 749, 142384.

Yin, Z. H., & Choi, C. H. (2023). Does digitalization contribute to lesser income inequality? Evidence from G20 countries. Information Technology for Development, 29(1), 61-82.

Zhang, C., Liu, B., & Yang, Y. (2024). Digital economy and urban innovation level: A quasi-natural experiment from the strategy of "Digital China". Humanities and Social Sciences Communications, 11(1), 1-12.

Zhang, J. (2021). A survey on income inequality in China. Journal of Economic Literature, 59(4), 1191-1239.