Inequality of Opportunity and Economic Growth: New Evidence

Francesco Colcerasa †

Department of Economics and Law, Sapienza University of Rome

April 30, 2023

This is a preliminary draft, please do not quote.

Abstract

This article estimates the effect of inequality of opportunity (IOp) on economic growth in European regions by providing an innovative analysis of such a nexus. Using Machine Learning to control for a high-dimensional vector of controls (i.e. human capital, institutions democracy, industrial structure, economic conditions, financial development, demographic aspects, urbanization, labour market conditions, investments, market distortions, public intervention) and mitigate potential endogeneity deriving from omitted confounders, we find a general negative effect of IOp on economic growth. This estimated negative effect persists even when controlling for total income inequality. Differently from previous empirical works, we also provide new evidence on the potential dynamic effect of IOp on economic growth by using local projections and observing our outcome variable over a four-years horizon. Interestingly, we find that the negative effect of IOp on economic growth is not persistent over time with statistically significant coefficients at the first two horizons only.

Keywords: Inequality of opportunity, Economic growth, European regions, Local projections, Machine Learning, Post-Double LASSO.

JEL Codes: D63, I30, 047, R11.

†Email address:francesco.colcerasa [at] uniroma1.it

1 Introduction

In principle, the issue of inequality can be approached from two main perspectives. The first one is moral-based, in which case it is discussed whether inequality is per-se ethically acceptable. Beyond the strong philosophical connotations, part of economic literature adopted such point of view discussing the pro-social preferences for inequality (Bolton and Ockenfels, 2000; Alesina et al., 2004; Ferrer-I-Carbonell and Ramos, 2014). Consistently with these theories, people may show different tastes for equality according with their cultural background. The second perspective approaches inequality issues by partially abandoning the moral argument to embrace an "agnostic" evaluation method, drawing from consequentialist philosophy (Anscombe, 1958). In particular, the goodness of inequality is evaluated by its effects rather than through purely ethical considerations. Accordingly, inequality is assessed positively (negatively) just in light of its positive (negative) effects on a third phenomenon (e.g. growth, well-being, political stability, etc.). Plenty of previous economic literature implicitly adopted such perspective, investigating the impact of inequality on a multitude of economic phenomena. Among these, an important strand of literature investigated whether inequality had an influence on economic growth. Nonetheless prosperous, such line of research has not reached yet a conclusive evidence about the existence and the possible direction of such link. In this field, several temporal phases can be identified. The first generation of studies focused on the role of income inequality in shaping growth. The second generation of studies – which the present work belongs to – adopted a different definition of inequality, namely inequality of opportunity.

Such methodological distinction was made to discuss the concept of fairness and to solve the inconclusiveness of the first generation about sign and significance of the relationship. Indeed, in that field, some authors find a positive link, emphasizing the role played by inequality – intended as concentration of resources in the hand of the richest – as a stimulus to investments and, in turn, to growth (Lewis, 1954; Kaldor, 1957; Pasinetti, 1962). These authors draw from Keynes (1936) conclusions about the distribution of the marginal propensities to consume and save across the income distribution. Contrary to the poorest, those at the top of the income ladder devote a higher share of their incomes to savings, while consuming a very small part of it. Therefore, consistently with the equilibrium equivalence of savings and investments, a higher portion of total income in the hands of those who save the most automatically translates in greater investments and growth. Among others, Li and Zou (1998) and Forbes (2000) provide the empirical counterpart supporting such theoretical view. However, such findings clash with other evidences advocating for a negative relationship between inequality and growth. In this regard, several channels transmitting a detrimental effect can be identified. In particular, Berg et al. (2014), in line with Alesina and Rodrik (1994) and Perotti (1993), emphasize that inequality may trigger stronger preferences for redistribution – formalized in an increased tax rate through the voting process – which imply a disincentive for investments, depressing growth.

At the same time, other studies, like those of Alesina and Perotti (1996) and Perotti (1996), highlight the channel represented by the political instability caused by inequality, which results in worse conditions for risky economic activities that are vital in the expansionary phases of the economy. Such perspective goes in hands with the idea that inequality may create unstable and fragile property rights as well as a fertile ground for criminality and lack of law enforcement, that further hinder development, as documented by Saint-Paul and Verdier (1993) and Persson and Tabellini (1994). All these factors are detrimental to growth, since a solid and stable socio-economic environment is vital for entrepreneurial activities, investments, and innovation (Barro, 2000). Alongside, other authors remark the distortionary role played by inequality in the process of human capital accumulation, which is crucial to support the economic development process and industrialization (Murphy et al., 1989; Galor and Zeira, 1993; Saint-Paul and Verdier, 1993; Galor and Moav, 2004).

The emerging picture is characterized by the presence of different mechanisms at work contemporaneously and pointing in different directions in mediating the inequality-growth nexus. As summarized in Voitchovsky (2012), many attempts to solve such ambiguity and inconclusiveness have been made. Some works, like those of Forbes (2000) and Banerjee and Duflo (2003), pinpoint model specification and econometric methodologies as the main causes of such contradictory patterns. Other authors argue that the choice of the measurement methodology of inequality affects the way it interacts with growth (Szèkeli, 2003). At the same time, some scholars hypothesize that the relationship between inequality and growth is sensitive to the data basis on which it is investigated (Deininger and Squire, 1996). Nonetheless pointing out the possible causes of such inconclusiveness contributed to make significant steps towards its solution, none of the provided explanations completely addresses it, leaving space for additional research.

In particular, the second generation of studies adopts a different perspective, shifting the focus of inequality from final outcomes (i.e, income) to its sources (i.e. opportunities). From a philosophical point of view, the juxtaposition between inequality of outcomes and inequality of opportunities roots back in the theory of distributive justice provided by Dworkin (1981). Its philosophical contribution consists in going upstream the distributive process and focus on the causes of unequal outcomes allocations. Drawing from such perspective, the seminal work of Roemer (1998) proposes the distinction between inequality of opportunity and inequality of effort, respectively stemming from factors beyond individual control and personal choices the individuals should be held accountable for. On the wake of such approach, some authors argue that when investigating the inequality-growth nexus, such distinction should be included in the discussion, supporting the idea that equality of opportunity matters for economic expansion, not income equality (Marrero and Rodriguez, 2013; Teyssier, 2013; Marrero et al., 2016; Bradbury and Triest, 2016; Ferreira et al., 2017; Carranza, 2020; Aiyar and Ebeke, 2020). Nevertheless, data scarcity severely limited the development of such research line.

The present work further develops this perspective, proposing a novel investigation of the relationship between inequality of opportunity and economic growth in European regions. In particular, it innovates previous literature in several aspects. First, it proposes a pioneering perspective about the relationship of interest, analysing the dynamic effect of inequality of opportunity on growth over time through local projections. In particular, it is tested how the link between inequality of opportunity and growth behaves across time. Second, it advances the empirical strategy refining as much as possible previous estimations of the relationship between inequality of opportunity and growth through machine-learning algorithms. Third, it proposes an empirical application to European data, where the relationship of interest has been investigated only in another work at the country level (Carranza, 2020). Fourth, it studies the effects of inequality of opportunity on growth at the regional level, taking into account sub-national heterogeneity while confronting different countries. Fifth, it improves the empirical estimation of the relationship by using novel regional measures of inequality of opportunity, based on an extended set of circumstances provided by EU-SILC minutely capturing parental background.

The present work is organized as follows. Section II reviews past literature about the nexus between inequality of opportunity and economic growth, emphasizing the research gaps it presents. Section III emphasizes the gaps the present work intends to fill, formalizing them in several research questions. Section VI describes how inequality of opportunity is obtained as well as the characteristics of the datasets and the variables of interest. Section V describes the empirical strategy adopted to estimate the impact of inequality of opportunity on growth. Section VI presents and comments the results Section VII provides some concluding remarks.

2 Literature Review

The potential link between inequality of opportunity (hencefort, IOp) and economic growth, although fascinating, is rather unexplored. Indeed, severe lack of data, both to accurately estimate inequality of opportunity and to relate it to growth, represent a major limitation in this field. Nevertheless, some works took this path and attempted to solve the inconclusiveness of previous findings, producing a variety of results discussed in the next paragraphs.

A first contribution is that of Marrero and Rodriguez (2013), who test the relationship in 23 US states in the 1980s and 1990s using the Panel Survey Income Dynamics (PSID). The empirical strategy proposed by the authors consists in taking two consecutive decades (1980-1990 and 1990-2000) and regressing growth rates in the ensuing 10 years on inequality of opportunity measured at the beginning of each decade. They include inequality of opportunity and income inequality jointly in the model, such that the effect of the latter is cleared by the impact of observed circumstances. Their empirical evidence shows pro-growth effects of income inequality and anti-growth effects of inequality of opportunity. However, their pioneering contribution relied on a non-parametric estimation of inequality of opportunity (Ferreira and Gignoux, 2011) employing only two circumstances, namely father's education and race. Their combination results in a low number of representative types on which inequality of opportunity is estimated. This may result in a lower variability of the representative median incomes on which inequality of opportunity is computed, affecting in turn the estimated relationship with growth. In other words, being based on a restricted number of circumstances, the true effect of unfair inequalities on growth may be perturbed.

A methodological extension of their work is proposed in Marrero et al. (2016), where they investigate the impact of inequalities on growth at various levels of the income ladder. Specifically, they test whether differences in the nexus exist across the income distribution. The estimation of inequality of opportunity is carried out on the IPUMS-USA dataset, which covers 50 years between 1960 and 2010. The econometric analysis is performed both through OLS and system-GMM estimation. The negative effect of inequality of opportunity on growth is confirmed, while the impact of total inequality (previously negative) becomes mostly insignificant. Re-examining the relationship restricted to different households according to their income class, a negative and positive relationship for poor and rich people respectively emerge. Accordingly, opportunities mis-allocation limit the income prospects of poorer rather than richer people. However, also in this case estimation methodology of inequality of opportunity follows a non-parametric approach, where only race and gender are used as circumstances. The small dimension of the circumstances' set jointly to the absence of socio-economic information about individual background (like parents' education or occupation, or socio-economic conditions when young) may undermine the estimation of inequality of opportunity. A similar scenario emerges from the methodological approach of Teyssier (2013), where socio-economic aspects are overlooked as well. However, in this case the empirical analysis is carried out in Brazil and leads to non-significant results.

Another notable contribution in this field is that of Ferreira et al. (2017). In this case, the inequality decomposition is performed on two meta-datasets, one formed by income and expenditure surveys and one formed by household demographic surveys. In the former case, a non-parametric approach is adopted (Ferreira and Gignoux, 2011) to partition the sample in circumstances' groups and apply a dispersion measure to the median income distribution across types. In the latter case, they regress a proxy of individual wealth[1] on a set of demographic circumstances, and take the variance of predicted values as a measure of inequality of opportunity. However, the socio-economic background of individuals is overlooked in the computation

[1] Obtained through a principal component analysis consistently with the methodology of Filmer and Pritchett (2001).

of inequality of opportunity. In fact, the circumstance set includes demographic aspects only - such as gender, race, language spoken at home, religion, immigration, region of birth - and does not consider the socio-economic background that is, instead, usually considered to play a major role in shaping the individual income prospects. The econometric analysis consists in regressing growth rates on the 5-year lagged index of inequality of opportunity, measured in either ways. Both pooled (OLS) and panel models (FE and GMM) are implemented. The emerging results about the sign and magnitude of the relationship are inconclusive, suggesting that the inequality decomposition does not solve the puzzle in growth literature. A similar empirical strategy, although with different results, is adopted by Carranza (2020). He investigates the role played by inequality of opportunity (measured in 2005 and 2011 only) in shaping the economic performance of 27 European countries, providing the unique investigation of such nexus on European data. He finds a negative and statistically significant effect of the former on economic growth.

Finally, Aiyar and Ebeke (2020) propose a different interpretation of the role of opportunities within the growth framework. First, they intend inequality of opportunity as intergenerational mobility, measured through the elasticity of individuals' income to parents' income or education. Second, they argue that mobility acts as a mediator of income inequality rather than as an alternative to it, as instead happens in the other studies. They hypothesize that the effect of income inequality on growth is mediated by society's rigidities, such that it is harmful for growth when intergenerational mobility is low and that it stimulates growth when mobility is high. They find significant results in support of their thesis, remarking the retarding effect of low mobility on growth. Similarly, Bradbury and Triest (2016) hypothesize the relevance of relative and absolute intergenerational mobility in shaping the growth path, supporting the explanation that inequality of opportunity retards growth because it encompasses a waste of resources. They test such explanation in a pooled regression, finding a significantly negative effect. However, when implementing an instrumental variable approach, the coefficient of intergenerational mobility becomes not significant.

Similarly to the first generation of studies, where the interaction between income inequality and growth was investigated, also in this sub-field a considerable ambiguity emerges. It relates to the conflict among findings, the variables underpinning the inequality decomposition, the lack of data, the adopted empirical strategies. The present work attempts to contribute to this research branch by addressing some of these research gaps. First, it attempts to solve the inequality-growth puzzle in terms of sign and significance using new estimates of inequality of opportunity for European regions. Second, a dynamic perspective is introduced, testing the impact of inequality of opportunity on local projections of growth. Third, it adopts an agnostic approach in regards of the model specification, using machine learning algorithms to address both endogeneity and model specification issues. Fourth, it relies on original estimates of inequality of opportunity at the regional level, obtained starting from a broad set of socio-economic and demographic circumstances enhancing the accuracy of the inequality decomposition[2]. Fifth, it provides a novel investigation of the role of inequality of opportunity on economic growth in taking sub-national territories as unit of interest.

3 Motivation: research gaps and research hypotheses

Previous attempts to disentangle the inequality-growth puzzle by considering opportunities allocation represent precious contributions moving forward the research horizon. However, several methodological facets leave space for further discussion. This section deepens such aspects, emphasizing the emerging research gaps and the additional contributions of the present article. Then, on such basis, it points out the research hypotheses to be tested in the empirical analysis, formalized in few synthetic research questions.

A first issue is represented by the absence of a dynamic investigation of the relationship between inequality of opportunity and economic growth. In particular, no previous work tested the evolution of their link – presumed that there is one – over time. Instead, previous literature focused on a static nexus between inequality of opportunity and growth averaged over subsequent years. A second problem stems from the lack of in-depth data matching individual incomes and parental background of earners. Its curbing role emerges for at least two reasons. First, from a methodological point of view, the quality of inequality decomposition depends on the number and type of available circumstances in the dataset. In principle, they should capture as much as possible all socio-economic and demographic factors in young age beyond individual control that may affect future income prospects. In this regard, previous research relies only on a restricted set of information about individuals' background when younger, giving space primarily to demographic variables and overlooking socio-economic factors, which instead are usually considered to play a prominent role in shaping future incomes. This may significantly dampen the inequality decomposition and, in turn, how it relates to growth[3]. Second, even though data about parental background are available, it is very rare that such information are reported consequentially for several years. This represents a major impediment to obtaining inequality estimates across time, crucial to perform growth regressions in a longitudinal framework. These aspects are particularly relevant for Europe, in which regards only one previous attempt at the country level has been made to investigate the relationship between inequality

[2]It is based on the three (2005, 2011, 2019) ad-hoc modules on intergenerational transmission of disadvantages provided by EU-SILC.

[3]In particular, as highlighted in Ferreira and Gignoux (2011), the partial inclusion of relevant circumstances downwards biases the estimation of inequality of opportunity. This is a standard issue when all "classic" circumstances are included, as many others are unobserved (e.g., the network of parents) and further exclusions may result in extremely an biased inequality decomposition.

of opportunity and growth (Carranza, 2020). Such issues are tightly related to other methodological vulnerabilities emerging from previous contributions. More specifically, each of the past studies provides a different model specification, such that it does not emerge a standard control set to include in growth regressions with inequalities. Indeed, past literature does not explicitly discusses which aspects should a growth regression account for while analysing the role of IOp, thereby affecting the comparability and robustness of past studies.

The present work addresses these issues through a novel empirical strategy estimating the nexus between inequality of opportunity and economic growth in European regions. Previous literature is innovated in several steps. First, a pioneering approach investigating the relationship of interest from a dynamic perspective is proposed. In particular, the effect of IOp on regional growth is tested through local projections of the latter over a time span of four years. Second, the econometric shortcomings stemming from ambiguous controls' sets and estimation methodology are mitigated, as much as possible consistently with data availability, in two steps. Firstly, a broad set of controls capturing potentially relevant aspects affecting growth is included, consistently with data availability. Secondly, since many control variables are included in the full model, dimensionality issues coming from model saturation may arise. Therefore, the model specification is regularized through machine learning, penalizing coefficients or ruling out regressors. In particular, a Post-Double LASSO estimation is implemented, addressing both dimensionality and endogeneity issues. This also helps to select an "optimal" set of controls, addressing the ambiguity displayed in past research. The rationale is to include all possible aspects scattered across previous works and let the algorithm select and use the most relevant ones. It is important to remark that the objective is not to determine a "one-fits-all" model specification, as it may vary according to different data, but let the algorithm select the best one consistently with the adopted data and consequently test the role of IOp. Third, as already emphasized, inequality decomposition is performed on a rich set of variables about individuals' parental background, much broader than those usually adopted in previous research. This contributes to reduce the bias in estimating inequality of opportunity. Fourth, the presence of multiple waves from the same dataset allows to investigate the relationship of interest in a panel framework. This allows for a more accurate estimation of inequality of opportunity and, consequently, for a better investigation of its effects on growth. Fifth, the analysis is performed on European regions, which have never been the object of any work in past literature. Then, a higher degree of territorial detail is deployed since the cross-sectional dimension is at the regional level. Such approach is original and has the merit of capturing within-country heterogeneities in growth patterns that may be overlooked in national comparisons. Indeed, the overall economic performance showed by a country may be an average of those of sub-national local economies going at different speeds and where the same growth drivers may weigh differently due to their peculiar characteristics (take, for instance, the case of northern and southern regions in Italy). All these contributions are formalized in the following research questions:

RQ_1: Is economic growth driven by inequality of opportunity in European regions?

RQ_2: Does total inequality still play a role once its circumstance-driven component is considered?

RQ_3: Is there a dynamic relationship over time between inequality of opportunity and economic growth

The research hypotheses these questions synthesize are tested in the subsequent sections, where a detailed discussion of data, empirical strategy and results is provided.

4 Data

The empirical analysis of the inequality-growth nexus relies on data coming from several sources. In particular, the dependent variable and the controls come from Eurostat, Penn World Tables, Polity V, while the main variable of interest – namely inequality of opportunity – is originally estimated on the EU-SILC data.

4.1 Inequality of Opportunity: estimation

In order to decompose total inequality in an ethically acceptable component (inequality of effort) and an ethically offensive component (inequality of opportunity), the present work adopts the parametric methodology provided by Ferreira and Gignoux (2011). The outcome of interest is household equivalised disposable income, as it is standard in field literature. Following the parametric approach, we first take the income distribution of each region in each year. Then, incomes are regressed over a vector of circumstances in an OLS framework with robust standard errors:

$$log(Y_i) = \alpha + \beta X_i + \epsilon_i \tag{1}$$

where X_i is the covariates vector, made of the set of variables capturing individual socio-economic background. Such set includes parents' education, parents' occupation, parents' activity status, parents' presence, household financial situation, immigration status, gender, number of siblings, urbanisation degree, for a total of 12 covariates. The estimated coefficients allow to obtain the predicted income, formalized as:

$$\hat{Y}_i = e^{(\hat{\beta} X_i)} e^{\frac{\sigma_\epsilon^2}{2}} \tag{2}$$

Importantly, the second exponential term is included to correct for the estimation bias implied by the reverse log-transformation, as suggested in Niehues and Peichl (2014) and Blackburn (2007) and Ferreira and Gignoux (2011). The distribution of predicted incomes is the counterfactual distribution, i.e. the one that roots back to the unfairness implied by individual socio-economic backgrounds. Therefore, applying a measure of inequality $I(\dot{)}$ allows for measuring inequality of opportunity, such that:

$$IOp_A = I(\hat{Y}) \tag{3}$$

which can also be expressed in relative terms with respect to total inequality:

$$IOp_R = \frac{I(\hat{Y})}{I(Y)} \tag{4}$$

The adopted inequality measure is the Gini index, which varies between 0 (perfect equality) and 1 (perfect inequality). The decomposition is entirely based on EU-SILC cross-sectional data, which are collected on annual basis in Europe and provide information at a regional level[4]. Moreover, each year a special module is issued to provide ad-hoc information about a particular topic. The present work is based on the waves issued in 2005, 2011, and 2019, where intergenerational transmissions of socio-economic disadvantages are investigated. In particular, individual data about childhood and parental background are collected. Such data are extremely suitable for the present analysis, as they allow to investigate the socio-economic roots from the past of subjects' achievement in the present. Descriptive statistics on household income and circumstance variables are presented in Table Table 1.

As is possible to note from Table 1, the estimation of inequality of opportunity relies on almost 700,000 observations, collected in 61 European regions across the 3 ad-hoc intergenerational modules of EU-SILC. The present work provides original estimates of inequality of opportunity, innovating previous findings from a multitude of perspectives. First, inequality of opportunity is estimated at the regional level, which contributes to shed light on European differences contemporaneously among regions of different states and among regions of the same state. Second, three waves of the EU-SILC are used (2005, 2011, 2019)[5], which enables to investigate the evolution of inequality components over an economically interesting period. Indeed,

[4]The regions are the following, consistently with EU-SILC classification and comparability over years (since the classification in some cases differs across waves): AT1, AT2, AT3, BE1, BE2, BE3, CZ01, CZ02, CZ03, CZ04, CZ05, CZ06, CZ07, CZ08, EL3, EL4, EL5, EL6, ES11, ES12, ES13, ES21, ES22, ES23, ES24, ES30, ES41, ES42, ES43, ES51, ES52, ES53, ES61, ES62, ES63, ES64, ES70, FI19, FI1B+FI1C+FI18, FI1D, FR1, FR2, FR3, FR4, FR5, FR6, FR7, FR8, HR0, HU1, HU2, HU3, ITC, ITF, ITG, ITH, ITI, PL1, PL2, PL3, PL4, PL5, PL6.

[5]Notice that EU-SILC data are backdated, such that the 2005, 2011, 2019 waves correspond, respectively, to information of the year 2004, 2010, 2018. The subsequent data are, then, matched accordingly.

in 2011 the public debt crisis arose, while in 2019 the pandemic still had to spread. Therefore, it is possible to analyse how inequality composition changed across three phases: the pre-crisis (2005), the crisis (2011), and the post-crisis (2019). Third, an extensive use of circumstances is made in obtaining IOp estimates to be related to growth. Previous research, to the best of the present knowledge, limited the analysis at most to parents' education and occupation, overlooking other aspects relevant to the socioeconomic characterization of individual backgrounds and, thus, to personal achievements.

Table 1: *Descriptive Statistics EU-SILC*

Variable	**Observation**	**Mean**	**Std. Dev.**	**Minimum**	**Maximum**
HH Equivalised Income	694745	16920.58	20071.38	.112	6178830
Gender	694745	.477	.499	0	1
Immigrant Status	694745	.095	.293	0	1
Birthplace Urbanization					
City	694745	.362	.481	0	1
Suburb	694745	.242	.428	0	1
Rural	694745	.328	.469	0	1
Number Children	694745	2.674	1.699	0	20
Parents Presence					
Both Parents	694745	.879	.325	0	1
Father Only	694745	.015	.122	0	1
Mother Only	694745	.089	.286	0	1
Private HH	694745	.013	.116	0	1
Collective HH	694745	.001	.032	0	1
HH Financial Situation					
Very Bad	694745	.056	.231	0	1
Bad	694745	.265	.441	0	1
Good	694745	.485	.499	0	1
Very Good	694745	.103	.304	0	1
Father's Education					
Low	694745	.526	.499	0	1
Medium	694745	.260	.439	0	1
High	694745	.105	.307	0	1
Mother's Education					
Low	694745	.618	.486	0	1
Medium	694745	.246	.431	0	1
High	694745	.075	.264	0	1
Father's Occupation					
Manager	694745	.056	.229	0	1
Professional	694745	.065	.246	0	1
Technician	694745	.075	.263	0	1
Clerical Support	694745	.042	.199	0	1
Services and Sales	694745	.057	.232	0	1

Skilled Agri-Fishery	694745	.118	.322	0	1
Craft and Trade	694745	.209	.406	0	1
Plant Operators	694745	.118	.323	0	1
Elementary	694745	.098	.298	0	1
Armed Force	694745	.011	.102	0	1
Mother's Occupation					
Manager	694745	.019	.138	0	1
Professional	694745	.057	.233	0	1
Technicians	694745	.052	.222	0	1
Clerical Support	694745	.065	.246	0	1
Services and Sales	694745	.093	.291	0	1
Agri-Fishery	694745	.084	.277	0	1
Craft and Trade	694745	.052	.222	0	1
Plant Operators	694745	.034	.181	0	1
Elementary	694745	.111	.314	0	1
Armed Force	694745	.001	.023	0	1
Father's Activity Status					
Employed	694745	.687	.463	0	1
Self-Employed	694745	.191	.393	0	1
Unemployed	694745	.006	.081	0	1
Retired	694745	.011	.103	0	1
Housework	694745	.002	.042	0	1
Other	694745	.013	.116	0	1
Mother's Activity Status					
Employed	694745	.457	.498	0	1
Self-Employed	694745	.109	.313	0	1
Unemployed	694745	.006	.075	0	1
Retired	694745	.007	.081	0	1
Housework	694745	.371	.483	0	1
Other	694745	.016	.126	0	1

Notes: Data are averaged on the three waves of EU-SILC used in this work. In the cases of the variables Parents Presence, Household Financial Situation, Parents' Education and Occupation, variables were harmonized across waves. In the case of Parents' Education, *Low* includes all categories below the lower secondary level, *Medium* includes all categories between the lower secondary level and the post-secondary-non-tertiary level, *High* includes all remaining categories up to doctoral level. In the case of Parents' Occupation, the ISCO-08 Classification was followed. In some cases, the categories' frequency do not sum to one. The missing share is due to lack of response, recoded in every case as a zero. This was done to preserve observations for the subsequent analysis.

Table 2: *Inequality Decomposition in European Regions*

European Regions	**2005** I_{TOT}	I_{OPP}	I_{EFF}	**2011** I_{TOT}	I_{OPP}	I_{EFF}	**2019** I_{TOT}	I_{OPP}	I_{EFF}
AT1	0,270	0,098	0,172	0,299	0,120	0,178	0,288	0,107	0,181
AT2	0,253	0,079	0,174	0,247	0,082	0,165	0,253	0,084	0,168
AT3	0,251	0,088	0,163	0,249	0,084	0,165	0,249	0,079	0,170
BE1	0,539	0,163	0,376	0,370	0,187	0,183	0,327	0,160	0,166
BE2	0,239	0,080	0,160	0,237	0,088	0,149	0,211	0,072	0,139

BE3	0,263	0,095	0,168	0,249	0,108	0,141	0,230	0,097	0,133
CZ01	0,283	0,101	0,182	0,289	0,088	0,201	0,262	0,087	0,175
CZ02	0,287	0,117	0,170	0,229	0,094	0,135	0,219	0,078	0,142
CZ03	0,219	0,080	0,139	0,224	0,075	0,149	0,210	0,067	0,143
CZ04	0,255	0,080	0,175	0,268	0,113	0,154	0,241	0,097	0,144
CZ05	0,235	0,080	0,155	0,247	0,085	0,163	0,205	0,073	0,132
CZ06	0,215	0,059	0,156	0,246	0,078	0,169	0,224	0,070	0,154
CZ07	0,241	0,085	0,157	0,246	0,086	0,160	0,230	0,081	0,149
CZ08	0,272	0,098	0,174	0,238	0,079	0,159	0,225	0,085	0,140
EL3	0,302	0,109	0,193	0,326	0,133	0,193	0,308	0,126	0,182
EL4	0,311	0,135	0,175	0,292	0,135	0,158	0,323	0,104	0,218
EL5	0,313	0,118	0,195	0,346	0,135	0,211	0,300	0,113	0,187
EL6	0,341	0,137	0,204	0,306	0,123	0,183	0,291	0,095	0,196
ES11	0,304	0,128	0,176	0,280	0,122	0,157	0,296	0,149	0,147
ES12	0,307	0,141	0,166	0,283	0,145	0,138	0,365	0,172	0,193
ES13	0,302	0,150	0,152	0,296	0,148	0,149	0,309	0,153	0,156
ES21	0,274	0,115	0,159	0,302	0,122	0,179	0,282	0,138	0,144
ES22	0,279	0,127	0,151	0,301	0,154	0,147	0,230	0,125	0,105
ES23	0,276	0,148	0,127	0,291	0,172	0,120	0,287	0,195	0,091
ES24	0,289	0,141	0,148	0,273	0,174	0,099	0,263	0,157	0,106
ES30	0,299	0,115	0,184	0,330	0,174	0,156	0,345	0,178	0,167
ES41	0,327	0,182	0,145	0,283	0,123	0,160	0,254	0,148	0,106
ES42	0,322	0,239	0,083	0,343	0,189	0,154	0,310	0,185	0,124
ES43	0,349	0,165	0,184	0,334	0,152	0,182	0,284	0,149	0,136
ES51	0,286	0,108	0,178	0,312	0,152	0,160	0,300	0,144	0,157
ES52	0,293	0,091	0,202	0,333	0,155	0,178	0,299	0,153	0,146
ES53	0,308	0,159	0,150	0,344	0,195	0,149	0,290	0,169	0,121
ES61	0,305	0,141	0,164	0,337	0,158	0,179	0,330	0,148	0,182
ES62	0,292	0,109	0,184	0,284	0,149	0,136	0,285	0,142	0,143
ES63	0,416	0,272	0,143	0,316	0,265	0,050	0,396	0,321	0,075
ES70	0,321	0,122	0,199	0,317	0,137	0,180	0,305	0,148	0,157
FI19	0,237	0,056	0,181	0,248	0,060	0,188	0,243	0,071	0,172
FI18	0,278	0,070	0,207	0,254	0,073	0,181	0,265	0,062	0,203
FI1D	0,248	0,062	0,186	0,252	0,073	0,179	0,241	0,065	0,175
FR1	0,285	0,137	0,148	0,319	0,134	0,185	0,325	0,151	0,174
FR2	0,246	0,088	0,158	0,274	0,090	0,184	0,254	0,094	0,161
FR3	0,280	0,125	0,155	0,278	0,126	0,152	0,263	0,142	0,121
FR4	0,246	0,086	0,160	0,293	0,088	0,205	0,262	0,103	0,159
FR5	0,234	0,075	0,159	0,246	0,079	0,167	0,251	0,058	0,193
FR6	0,297	0,076	0,221	0,304	0,092	0,212	0,273	0,080	0,193
FR7	0,264	0,091	0,173	0,293	0,096	0,197	0,240	0,088	0,152
FR8	0,283	0,096	0,186	0,307	0,110	0,197	0,267	0,103	0,164
HU1	0,292	0,090	0,202	0,282	0,134	0,148	0,334	0,124	0,210
HU2	0,260	0,088	0,172	0,251	0,100	0,151	0,267	0,104	0,163
HU3	0,269	0,098	0,171	0,269	0,117	0,152	0,255	0,112	0,143
ITC	0,303	0,102	0,201	0,289	0,105	0,184	0,300	0,111	0,189
ITF	0,326	0,108	0,218	0,327	0,126	0,201	0,331	0,126	0,205
ITH	0,284	0,084	0,200	0,282	0,095	0,187	0,261	0,098	0,163
ITI	0,295	0,087	0,208	0,299	0,117	0,182	0,290	0,113	0,178
ITG	0,343	0,149	0,194	0,339	0,137	0,202	0,346	0,138	0,208
PL1	0,409	0,186	0,223	0,347	0,159	0,187	0,325	0,153	0,172
PL2	0,335	0,110	0,225	0,315	0,111	0,204	0,276	0,101	0,176
PL3	0,340	0,142	0,197	0,312	0,137	0,175	0,282	0,129	0,153
PL4	0,350	0,122	0,228	0,295	0,127	0,168	0,268	0,112	0,156
PL5	0,362	0,123	0,240	0,317	0,113	0,204	0,278	0,109	0,169
PL6	0,350	0,145	0,205	0,310	0,134	0,176	0,289	0,119	0,170

As is possible to note from Table 2, inequality - regardless of its type - behaved in different ways among European regions across the years. In some cases, like in a Northern-Centre of France (FR1) or in the Southern Italy (ITF), a strong uprise in inequality was experienced. Such rise, was accompanied by a parallel increase in inequality of opportunity. In other cases, like in Central Hungary (HU1), or Eastern Austria (AT1) total inequality and inequality of opportunity went in opposite directions, reaching a minimum and a peak, respectively, in 2011. However, both types of inequality reached a higher level in 2019 than in 2005. Other regions, such as the Attic area in Greece (EL3) or Basque Spain (ES21), experienced a stable increase in every type of inequality, although with a peak in 2011. Others went in the opposite direction, like in most of Poland. The causes of such mixed patterns are manifold, and could vary across countries and regions. Surely, a historical component plays a role. In every country, some regions have historical socio-economic difficulties which do affect their capability to cope with the crises. Exemplary is the case of Southern Italy, where every type of inequality increased across the considered time span.

4.2 Additional data

The empirical analysis relies on variables collected at the regional level, following the NUTS classification and consistently with the EU-SILC nomenclature[6]. All regions are followed for a period of 18 years, between 2004 and 2022. Growth rates are computed as the percentage change of real GDP per-capita between consecutive years. The dependent variable is the average of these growth rates over a time span of four years subsequent to each baseline period, namely the moment in which inequality of opportunity is observed. The control variables are taken contemporaneously to the main variable of interest, as is standard in field literature. Although the cross-sectional unit of observation is the regional level, some controls are only unfortunately available at the national level only. Therefore, the set of covariates is composed by variables with different territorial declinations[7]. Such data come from different sources, accordingly with

[6]The regional classification of control variables must be consistent with the EU-SILC classification and comparability over years (since the classification in some cases differs across waves) because it drives the way inequality of opportunity is collected. The regions are the following: AT1, AT2, AT3, BE1, BE2, BE3, CZ01, CZ02, CZ03, CZ04, CZ05, CZ06, CZ07, CZ08, EL3, EL4, EL5, EL6, ES11, ES12, ES13, ES21, ES22, ES23, ES24, ES30, ES41, ES42, ES43, ES51, ES52, ES53, ES61, ES62, ES63, ES64, ES70, FI19, FI1B+FI1C+FI18, FI1D, FR1, FR2, FR3, FR4, FR5, FR6, FR7, FR8, HR0, HU1, HU2, HU3, ITC, ITF, ITG, ITH, ITI, PL1, PL2, PL3, PL4, PL5, PL6.

[7]In particular, data about productivity, financial development, international trade, market distortions, institutions' democracy, government expenditure are taken at the country level.

their availability[8]. The choice of which aspects to control for is consistent and covers all of the previous field literature (Marrero and Rodriguez, 2013; Aiyar and Ebeke, 2020; Ferreira et al., 2017; Carranza, 2020). In particular, we include demographic factors, human capital, employment, investments, institutions' democracy, industrial mix, financial development, public intervention. In particular, the variables used are real GDP per-capita, population level, fertility rate, mortality rate, old-age dependency ratio, urbanization, human capital, activity rate, employment rate, investment, sectoral composition (GVA), real productivity[9], financial development, terms of international trade, market distortions, institutions' democracy, government intervention. A legend explaining more in depth each variable can be found in Table 3. Instead, Table 4 provides the descriptive statistics of each variable.

Table 3: *Variables' Legend*

Variable	**Legend**
Dependent Variable	
Growth	It consists in the average growth rate over a span of 4 years, subsequent to the baseline period, of real GDP per-capita. It is available at the regional level.
Inequality Variables	
Total Inequality	It is the Gini index computed over observed household equivalised disposable incomes. It is available at the regional level.
Inequality of Opportunity	It is the Gini index computed over household equivalised disposable incomes predicted on the basis of individual circumstances. It is available at the regional level.
Controls	
Real GDP Per-Capita	It is computed as the ratio between real GDP and the population in the region. It is available at the regional level.
Population	It is the number of people living in the country. It is available at the regional level.
Fertility Rate	It is the mean number of children that would be born alive to a woman during her lifetime if she were to pass through her childbearing years conforming to the fertility rates by age of a given year. It is available at the regional level.
Mortality Rate	It is the ratio between the number of deaths and the population level. It is available at the regional level.
Old-Age Dependency Ratio	It is an Eurostat indicator of a population age structure. In particular, it is computed as the ratio between the population aged over 65 years old and the population aged between 20 and 64 years old. It is taken at the regional level.
Urbanization	It is the share of people living in rural areas and small villages. It is available at the regional level.
Human Capital	It is the share of people in the age brackets 25-64 years old who attained at least a lower tertiary education level (ISCED 7-8). It is taken at the regional level.

[8] All variables are provided or obtained starting from Eurostat data, except for the terms of the international trade and market distortions (provided by the Penn World Tables database) and the institutions' democracy (provided by the POLITY database).

[9] Real productivity is reported as an index referring to 2015 as baseline year, with a value of 100.

Activity Rate	It is the share of active people, namely the ratio between those belonging to the labour force and the population level. It is available at the regional level.
Employment	It is the share of employed people belonging to the labour force. It is available at the regional level.
Investment	It consists in the amount of investments in real terms. In particular, it is computed as the total of acquisitions, net of disposals, of assets intended for use in production processes (for instance, buildings, structures, machinery and equipment, etc.). It is included at the sectoral level (namely, agricultural, industrial, construction, services, financial, other), and it is taken as in ratio to GDP. It is available at the regional level.
Gross Value Added	It is defined as output value at basic prices less intermediate consumption valued at purchasers' prices, and it is calculated before consumption of fixed capital. It is included at the sectoral level (namely agricultural, industrial, construction, services, financial, other), and it is taken as in ratio to GDP. It is available at the regional level.
Real Productivity	It is an Eurostat indicator capturing the labour productivity per person in real terms. It is available at the country level.
Financial Development	It is and indicator provided by Eurostat, computed as the ratio between the amount of private credit and the GDP. It is available at the country level.
Terms of International Trade	It is the ratio between export prices and import prices, provided in the Penn World Tables database (PWT). It is available at the country level.
Market Distortion	It is and indicator provided by the Penn World Tables database (PWT). In particular, it is computed as the ratio between investments price level at purchasing parity to that at market prices. It is available at the country level.
Institutions' Democracy	It is an indicator provided by the POLITY database. It is the sum of two indicators, one capturing how much a country is autocratic - ranging between -10 and 0, and another one capturing how much a country is democratic - ranging between 0 and 10. It is available at the country level.
Government Total Expenditure	It is the ratio between total public expenditure and GDP. It is available at the country level.
Government Consumption Expenditure	It is the ratio between public expenditure in consumption (captured by education and military expenses) and GDP. It is available at the country level.
Government Social Expenditure	It is the ratio between public expenditure in social welfare, as defined by Eurostat, and GDP. It is available at the country level.

Source: Eurostat, POLITY, PWT, author's calculations.

Table 4: *Variables' Legend*

Variable	**Obs.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Regional Growth Rate	183	.010	.018	-.045	.067
Inequality of Opportunity (GINI)	183	.117	.039	.054	.321
Income Inequality (GINI)	183	.286	.039	.205	.416
Real GDP per-capita	183	23795.9	11668.3	5498.3	70936.1
Population	183	4296.19	3651.63	71.6	16078.8
Fertility rate	183	1.477	.268	.93	2.11
Mortality rate	183	9.257	2.191	.042	14.191
Old-Age Dependency Ratio	183	.289	.059	.179	.432
Urbanization	183	.363	.227	0	1

Human Capital	183	.216	.090	.058	.474
Activity Rate	183	.753	.061	.565	.865
Employment Rate	183	.714	.054	.533	.826
Investment to GDP Ratio (agr.)	183	.007	.005	-.002	.027
Investment to GDP Ratio (ind.)	183	.048	.028	.007	.183
Investment to GDP Ratio (con.)	183	.009	.008	.001	.067
Investment to GDP Ratio (ser.)	183	.039	.014	.015	.105
Investment to GDP Ratio (fin.)	183	.078	.028	.018	.139
Investment to GDP Ratio (oth.)	183	.038	.018	.012	.178
GVA Growth Rate	183	.025	.028	-.064	.120
GVA to GDP Ratio (agr.)	183	.028	.021	.001	.093
GVA to GDP Ratio (ind.)	183	.193	.080	.039	.393
GVA to GDP Ratio (con.)	183	.060	.019	.013	.115
GVA to GDP Ratio (ser.)	183	.219	.049	.134	.368
GVA to GDP Ratio (fin.)	183	.191	.055	.089	.381
GVA to GDP Ratio (oth.)	183	.201	.049	.109	479
Real Labour Productivity	183	97.323	8.007	76.749	116.255
Private Credit to GDP Ratio	183	.052	.057	-.039	.194
Terms of Trade	183	1.065	.038	.987	1.187
Market Distortions	183	.697	.126	.469	1.001
Institutions' Democracy	183	9.847	.443	8	10
Government Expenditure to GDP Ratio (tot.)	183	.467	.054	.388	.569
Government Expenditure to GDP Ratio (con.)	183	.061	.009	.048	.080
Government Expenditure to GDP Ratio (soc.)	183	.174	.036	.119	.243

Source: Eurostat, POLITY, PWT, authors' elaborations.
Notes: population is expressed in thousand millions, mortality rate is the number of deaths per thousands people.

5 Empirical Strategy

The empirical analysis aims at investigating the relationship between economic growth and inequality of opportunity, attempting to address the aspects discussed in previous paragraphs, within the limits imposed by data availability and theoretical assumptions. The last section remarked the improvements implied by the current inequality decomposition, whose outcomes are used as main covariate of interest in growth regressions. Concerning the empirical debate, previous works analysed the relationship in both cross-sectional (POLS) and panel frameworks (FE and GMM). However, beyond the well-known issues of cross-sectional estimation in growth frameworks, a recent contribution by Chen et al. (2019) remarks that FE and GMM models tend to be biased when the number of variables is high relatively to the sample dimension. In fact, the present analysis includes all aspects fragmentarily proposed in literature as controls in growth regressions. This considerably increases the number of covariates and may dampen the estimation efficiency. Indeed, the adoption of an ample control set is likely to generate model saturation, namely an increase in the model fitting despite a lower prediction power of several covariates Belloni et al. (2014). In order to avoid such shortcomings, machine learning techniques are implemented to regularize the model and mitigate these issues. This also contributes to address model selection questions, arising from the different and contrasting choices made in previous works about which aspects have to be controlled for when analysing growth determinants. Moreover, it helps to solve the issue of dimensionality and ensures that relevant regressors are included, addressing potential endogeneity shortcomings stemming from omitted variable bias. In particular, a Post-Double LASSO procedure is performed to select only those covariates that are relevant for growth and/or for inequality of opportunity. In other words, benchmark estimates are replicated exploiting machine learning techniques selecting the best predictors of both growth and inequality of opportunity. Such approach contributes to several objectives. Firstly, it regularize the model specification, correcting for model saturation and balancing the prediction ability and efficiency. Secondly, it accounts for endogeneity, avoiding to opt out poor predictors of growth which are endogenously correlated with inequality of opportunity, whose exclusion would generate an omitted variable bias. Thirdly, it helps to control for all growth-relevant aspects, in which regard no definitive position emerges from previous literature.

The empirical investigation starts from a baseline analysis in a panel framework, where observation-specific and time-specific effects are accounted for through a Two-Way Mundlak approach (Mundlak, 1978; Wooldridge, 2021). Such technique is particularly suitable to the present empirical strategy because it allows to simulate a two-way fixed effect model in a cross-sectional framework. This considerably simplifies the computational burden when implementing machine learning routines. A very large control set is adopted to capture all growth-relevant dimensions fragmentarily proposed in previous work. However, the inclusion of an ample set of variables is likely to generate model saturation, namely an increase in the model fitting although a lower prediction power of several covariates (Belloni et al., 2014). Being the time-specific and unit-specific heterogeneities stringent, it is necessary to control for such aspects. This is done by including cross-sectional and longitudinal averages of the control set, accounting for the unobserved heterogeneities. This procedure gives a benefit related to the subsequent machine learning estimation. In fact, in a FE framework the cross-validation routine in the Post-Double LASSO approach should create folds keeping into account the fixed effects. This would considerably increase the level of complexity and computation intensity of the routine. In fact, each fold should could not be created without placing cross-sectional units in different times in the same group. Instead, the Two-Way Mundlak estimation captures these heterogeneities while keeping a cross-sectional structure of the model.

5.1 Two-Way Mundlak Model

The Two-Way Mundlak follows the econometric methodology developed in Wooldridge (2021) in the wake of the work of Mundlak (1978). It consists in simulating a two-way fixed effect model, where both unit-specific and time-specific heterogeneities are accounted for, by estimating a RE model that includes both longitudinal and cross-sectional averages of all control variables. Such procedure reveals to be particularly useful. On one hand, it enables to account for heterogeneities otherwise overlooked in simple POLS or RE, and, on the other hand, it allows to keep a flexible model structure in view of the subsequent analysis encompassing machine learning algorithms.

In the benchmark framework, the empirical strategy proceeds as follows. First, a baseline model including an extended set of controls (and the related cross-sectional and longitudinal averages) is specified. The motivation is twofold. On one hand, the inclusion of many controls allows to take into account at the same time each of the aspects fragmentarily enclosed in each of previous works. This allows to check the presence of an interaction between growth and any type of inequality, while all possible confounding factors hypothesized in literature are controlled for. On the other hand, an extended set of covariates necessarily saturates the model, attenuating the bias but increasing the variability (Belloni et al., 2014), paving the path towards machine learning regularization, that is the object of the next subsection. The baseline model is specified as follows:

$$\bar{GR}_{r,t+4} = \alpha_0 + \beta_1 I_{r,t} + \theta_1 RegCon_{r,t} + \theta_2 NatCon_{c,t} + \omega_1 AvgS_r + \omega_2 AvgT_t + \epsilon_{r,t+4} \tag{5}$$

where the dependent variable ($\bar{GR}$) is the average growth rate of real GDP per-capita at the regional level computed over a period of 4 years subsequent to the year in which inequality of opportunity is measured. This strategy is standard in previous literature and helps to rule out endogeneity due to reverse causality between IOp and growth. The vector I includes both inequality of opportunity (IOp) and the effect of residual inequality, captured by total income inequality (TI). The former is the main variable of interest, measuring the "unfair" component of the income distribution at the regional level, consistently with the ethical formulation of Dworkin (1981) and Roemer (1998). A first specification is estimated only by including IOp. However, it is possible that its coefficient partially incorporates the effect of observed income inequality, therefore being biased. Consequently, in order to test the robustness of IOp, a model with only TI and one including both indicators are estimated. In the fully specified case, the coefficient of IOp is cleared for possible confounding effects of the residual inequality component captured by TI. Regarding the control set, although the analysis is performed at the regional level, some variables - such as the ratio of exports and imports price levels - are only available at the country level. In the vector $RegCon$ are included the following regional level covariates: logarithm of real GDP per-capita, population level, fertility rate, mortality rate, old-age dependency ratio, urbanization, human capital, activity rate, employment rate, investment to GDP ratio (by sector), GVA to GDP ratio (by sector). In the vector $NatCon$ are included the following country level covariates: real productivity, financial development, terms of international trade, market distortions, institutions' democracy, government expenditure (total, consumption, and welfare). The vector $AvgS$ includes the cross-sectional averages of the covariates, while the vector $AvgT$ includes the longitudinal averages of the covariates. In this framework, they configure as additional regressors in the model. Such control set, conditional on data availability, covers all aspects enclosed separately in each of those employed in empirical growth models investigating the impact of IOp. In a further specification, it is tested whether a dynamic relationship between IOp and growth exists. In particular, it is hypothesized that the effects of IOp fades away over time. In order to check such hypothesis, the following model specification is estimated:

$$GR_{r,t+j} = \alpha_0 + \beta_1 I_{r,t} + \theta_1 RegCon_{r,t} + \theta_2 NatCon_{c,t} + \omega_1 AvgS_r + \omega_2 AvgT_t + \epsilon_{r,t+j} \tag{6}$$

where *j=[1,4]*. In other words, the local projections of growth in each of the subsequent four years (over which the baseline growth rates are computed) are separately regressed over the same model specification of Equation 5. However, in both the static and the dynamic specification, the presence of a high number of controls is likely to saturate the model, generating a trade-off between estimation accuracy and efficiency. In this case, the application of machine learning techniques may help to select the best predictors of growth among the covariates in the control set, while accounting for possible bias due to their endogeneity in relation to inequality of opportunity.

5.2 Post-Double LASSO Model

In principle, unless a structural identification of the covariates is available, model specification is made on a slippery ground since irrelevant variables may be included and relevant ones may be ruled out. This is particularly true in the branch of inequality and growth. More specifically, model selection arises especially when the focus is on inequality decomposition, due to the novelty of such sub-field and to the non-unanimous choices made in past literature. Indeed, several studies estimated the impact of inequality of opportunity on growth, each considering as relevant different aspects (e.g. some exclude employment, others exclude market distortions, etc.). Moreover, being the present analysis a regional investigation of the local economic growth, several factors at the national level may be overlooked. In general, model specification failing to control for valid regressors may lead to biased parameters' estimation and unreliable findings. An immediate response to such issues may be to include as many controls as possible in the model structure. However, as emphasized in Belloni et al. (2014) and Belloni et al. (2017), the inclusion of a high number of controls may equivalently curb the analysis, due to model dimensionality and to spurious covariates' selection. In fact, the closer the number of controls and the sample dimension the higher the model saturation, which results in an increase of the standard errors. A trade-off between prediction accuracy and efficiency arises, calling for the exclusion of some regressors. Although guided by theory, in many cases the choice of which covariates have to be excluded from the specification may still encompass a certain degree of arbitrariness. In this regard, machine learning emerges as a helpful and effective approach to implement model regularization and select an optimal control set, balancing relevance and accuracy in estimation while remaining agnostic on which covariates should be excluded. In particular, according to the specific technique, machine learning adds helps to shrink and/or penalize each covariate's coefficient by testing its prediction power on data through cross-validation. The combination of shrinkage/penalty parameters that maximizes the prediction power of the model gives an associated control set that will be selected, ensuring model efficiency and accuracy.

In addition, when the empirical analysis aims at understanding the role of a certain regressor on the dependent variable, also possible endogeneity must be addressed. In fact, a classical machine learning LASSO routine may rule out a covariate with poor prediction power in respect of the dependent variable, but highly correlated to the regressor of interest. In that case, endogeneity from omitted variable bias would emerge. Therefore, the adopted machine learning technique should contemporaneously address prediction and endogeneity issues. In this regard, the Post-Double LASSO approach is particularly suitable to effectively estimate the impact of the variable of interest while correctly specifying the model[10]. Indeed, a classical LASSO estimator only excludes/shrinks poor predictors of the dependent variable. Hence, it is possible that the LASSO methodology would drop some covariates being weak predictors although strongly correlated with the regressor of interest. In such case, the coefficient of such regressor would be biased because of omitted variable endogeneity. This shortcoming can be tackled through the Post-Double LASSO technique, that identifies the regressors to be included in two steps. First, the good predictors of the dependent variable are selected, regardless their correlation with the regressor of interest. Second, those correlated to the covariate of interest are selected, regardless their prediction ability on the dependent variable. Then union of the two sets of regressors stands out as the optimal control set, addressing both regularization and endogeneity issues, which can be used in a final regression analysis to analyse the causal effect of the regressor of interest.

Therefore, in the present analysis, two different LASSO regressions are estimated. In the first one, the growth rate is regressed over the extended set of controls saturating the model:

$$\bar{GR}_{r,t+4} = c_0 + \tau_1 RegCon_{r,t} + \tau_2 NatCon_{c,t} + \delta_1 AvgS_r + \delta_2 AvgT_t + e_{r,t+4} \tag{7}$$

[10] For more technical details, see Urminky et al. (2016) or Belloni et al. (2014)

Overall, 93 regressors (comprehending the cross-sectional and the longitudinal averages) are included in the control set. A second regularization procedure is performed to account for endogeneity on the same set of controls but using IOp as a dependent variable:

$$IOp_{r,t} = d_0 + \gamma_1 RegCon_{r,t} + \gamma_2 NatCon_{c,t} + \eta_1 AvgS_r + \eta_2 AvgT_t + u_{r,t} \tag{8}$$

Applying a second specular regularization procedure enables to detect the presence of regressors that are poor predictors of growth - that would then be ruled out in the first LASSO - but highly correlated to inequality of opportunity, such that their exclusion would cause the endogeneity of the variable of interest due to an omitted variable bias. Therefore, the final specification will include a group of controls resulting from the union of the two selected set:

$$\bar{GR}_{r,t+4} = f_0 + \psi_1 I_{r,t} + \phi_1 selected_1 + \phi_2 selected_2 + h_{r,t+4} \tag{9}$$

which is estimated through both OLS and RE models with clustered standard errors at the regional level. In order to test for the robustness of IOp, the specification is estimated both with and without including total inequality in the vector I. Indeed, as mentioned above, it may happen that the effect of IOp may actually partially or totally absorb that of residual inequality. Moreover, the same machine learning procedure is repeated by keeping fixed IOp and using total income inequality as the variable of interest in the Post-Double LASSO routine.

The same procedure is applied to test for the presence of a dynamic relationship between IOp and growth. In particular, the same routine is repeated using the local projections of growth in each year of the period over the baseline growth rates are computed. Hence, the same model specification of Equation 7 will be estimated, but the dependent variable will be $GR_{r,t+j}$ instead of $\bar{GR}_{r,t+4}$, with *j=[1,4]*. Therefore, four additional Post-Double LASSO estimation are performed, one for each of the local projections.

6 Empirical Findings

The starting hypothesis of the present work is that the contrasting evidence in growth literature about the effect of inequality can be explained by decomposing inequality of outcomes and including the circumstance-driven share in the empirical analysis. The emerging findings from European data at the regional level do not reject such hypothesis, advocating for a negative and statistically significant effect of inequality of opportunity on the average growth rate of real GDP per-capita.

The baseline results are reported in Table 5, where the relationship between inequality of opportunity and growth is estimated in the Two-Way Mundlak Model (Mundlak, 1978; Wooldridge, 2021), with clustered standard errors at the regional level. Each column of Table 5 presents a different model specification, following a well-defined rationale. First of all, it is tested the role of income inequality alone (column 1), following the first generation of studies about the relationship between income distribution and growth. In this case, it emerges a negative effect, although non-statistically significant. Then, the subsequent analysis tests the hypothesis that isolating the circumstance-driven component of inequality of outcomes may solve the puzzle. The role of inequality of opportunity is tested in column 2, where a negative and statistically significant effect emerges. Still, part of its impact may be due to a confounding role of residual inequality. Therefore, in order to disentangle the effect of different types of inequality on growth, the model estimates presented in column 3 also include income inequality as additional control - capturing the role of residual inequality on growth. The sign of IOp resists this further robustness, remaining negative and statistically significant, while total inequality is positive but not significant. This preliminary evidence suggests that inequality of opportunity, not income inequality, matters in driving growth. In particular, the fully specified estimation suggests that a one-point increase in the Gini index of inequality of opportunity decreases the average growth rate across the subsequent four years of 0.076 p.p., while the effect of an equivalent increase in total inequality is not statistically different from zero.

Table 5: *Regression Analysis - Two-Way Mundlak Estimation*

	(1)	(2)	(3)
VARIABLES	RE	RE	RE
Inequality of Opportunity		-0.059*	-0.076**
		(0.031)	(0.035)
Income Inequality	0.010		0.037
	(0.037)		(0.038)
Controls	Yes	Yes	Yes
Cross-Sectional Averages	Yes	Yes	Yes
Longitudinal Averages	Yes	Yes	Yes
Observations	183	183	183
N. Regions	61	61	61

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

However, as emphasized above, the inclusion of many controls to cover all model specifications proposed in literature may lead to model saturation, triggering a trade-off between model efficiency and prediction accuracy. These shortcomings may lead to biased and unreliable results, calling for some regularization procedures. In this regards, machine learning techniques are particularly suitable tools to address the issues arising from model dimensionality. More specifically, in the present case a Post-Double LASSO algorithm is implemented to select and/or shrink the parameters of the model while tackling endogeneity. In fact, a classical LASSO procedure may opt out regressors that are poor predictors of the dependent variable (i.e. growth), but highly correlated to the covariate of interest (i.e. IOp), generating an omitted variable bias. Instead, the implementation of two LASSO routines avoids such issues, selecting the optimal predictors of growth and, at the same time, avoiding omitted-variable bias sourcing from the exclusion of correlated covariates with IOp. Therefore, the final model controls for selected regressors only, following the specification of equation Equation 9. Empirical results are presented in Table 6, where the definitive specification is estimated both including and excluding total inequality, either through OLS or RE.

Table 6: *Regression Analysis - Post-Double LASSO Estimation (TI fixed)*

Variables	(1) OLS	(2) OLS	(3) RE	(4) RE
IOp	-0.053*	-0.061**	-0.056*	-0.072**
	(0.029)	(0.030)	(0.031)	(0.034)
TI		0.019		0.035
		(0.037)		(0.037)
Controls	Yes	Yes	Yes	Yes
Cross-Sectional Averages	Yes	Yes	Yes	Yes
Longitudinal Averages	Yes	Yes	Yes	Yes
Observations	183	183	183	183
N. Regions	61	61	61	61
R-squared	0.935	0.935		

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In column 1 is reported the resulting evidence from a specification estimated through OLS where total inequality is excluded. In this case, a negative and statistically significant effect emerges. Such result is robust to the inclusion of total inequality in the model to control for possible effects of residual inequality. Only a marginal variation in the standard error of IOp coefficient is reported. In both cases, inequality of opportunity has a negative and statistically significant effect on the four-year average growth rate. More specifically, a one point increase in the Gini index of IOp results in a 0.061 p.p. decrease in the average growth rate in the subsequent four years. The same finding emerges when the same specification is estimated in a RE framework. In particular, column 3 is specular to column 1, with a statistically significant and negative, although slightly stronger, effect of IOp. In this case, the inclusion of residual inequality (column 4) makes the coefficient slightly lower. However, it does not affect the conclusion arising from the first two columns, suggesting that the true driver of growth is the type of inequality deriving from circumstances beyond individual control. In fact, the coefficient of residual inequality is negative although not statistically significant, while a one point increase in the Gini index of IOp implies a 0.072 p.p. decrease in the average growth rate.

As robustness check, an additional Post-Double LASSO routine is performed, using total inequality as variable of interest and keeping IOp fixed in the final specification. In this case, the second LASSO procedure is performed using income inequality as dependent variable, while IOp enters the model only in the final specification after covariates have already been selected. These results are presented in Table 7, whose pattern resembles previous table.

In particular, column 1 provides the evidence emerging from the final model specification, estimated through OLS, where only total inequality is included. A negative, but not statistically significant effect emerges. Instead, when also IOp is included, the coefficient of total (residual) inequality shrinks almost to zero, while the effect of the circumstance-driven component is negative and statistically significant. More specifically, a one-point increase in the Gini index of IOp leads to a 0.057 p.p. decrease in the average growth rate. A similar picture is depicted when the model is estimated in a RE framework. More specifically, when only total inequality is included (column 3), a negative but not statistically significant coefficient arises. When IOp is included as a covariate, the coefficient of total inequality turns positive, although very close to 0, and still not significant. Instead, the coefficient of IOp is negative, statistically significant, and slightly lower than that reported in column 3. In particular, a one-point increase in the Gini index of IOp leads to a 0.066 p.p. decrease in the subsequent average growth rate.

The emerging evidence suggests that a possible explanation to previous contrasting findings in literature can be found by isolating the "unfair component" from total inequality. In particular, a robust negative effect of inequality of opportunity is reported in both the Two-Way Mundlak estimation and the Post-Double LASSO routines. In particular, taking the last column of Table 6 it is possible to exemplify the relationship between IOp and growth as follows.

Considering a country whose Gini index and average growth rate are, for instance, 0.27 and 1.572%, an increase of the former to 0.28 leads the latter to fall to 1.5%. However, such relationship is static, namely captures only the effect of IOp on growth when the latter is averaged over a subsequent time span of multiple years.

Table 7: *Regression Analysis - Post-Double LASSO Estimation (IOp fixed)*

Variables	(1) OLS	(2) OLS	(3) RE	(4) RE
IOp		-0.057*		-0.066**
		(0.029)		(0.032)
TI	-0.023	-0.003	-0.018	0.006
	(0.031)	(0.031)	(0.031)	(0.031)
Controls	Yes	Yes	Yes	Yes
Cross-Sectional Averages	Yes	Yes	Yes	Yes
Longitudinal Averages	Yes	Yes	Yes	Yes
Observations	183	183	183	183
N. Regions	61	61	61	61
R-squared	0.931	0.933		

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.1 Local Projections

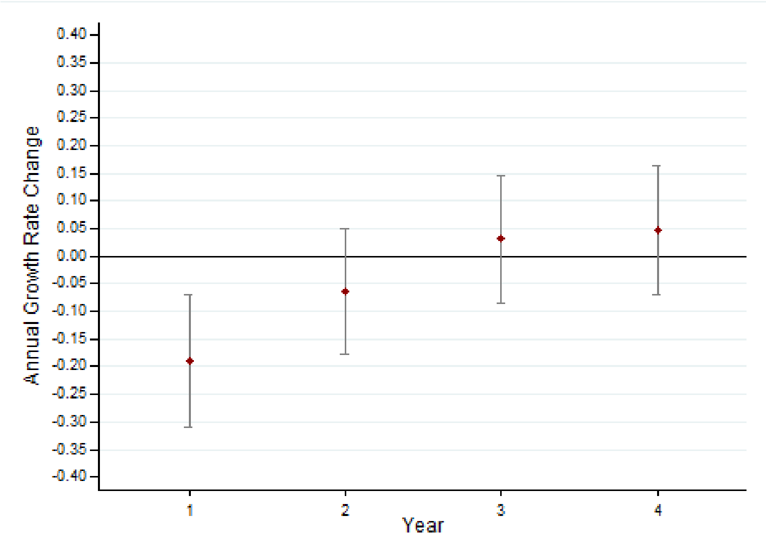
The previous section establishes a negative link between inequality of opportunity and economic growth. However, such effect is only an average over a medium-term period of 4 years. One may wonder whether there is a dynamic relationship between IOp observed in the baseline period and the punctual growth rate, taken over the same period of time of the average one used in the benchmark analysis. This an additional innovation the present work proposes. More specifically, the previous analysis is repeated using the annual growth rate in each of the subsequent four years as dependent variable (local projections), using both the Two-Way Mundlak model and the Post-Double LASSO routine for every local projection. The same approach as before is followed, providing evidence both excluding and including residual inequality.

In Figure 1(a) it is showed the dynamic evolution of IOp coefficient estimated through the Two-Way Mundlak model, when residual inequality is not accounted for. It emerges a dynamic negative effects fading away over time. More specifically, the negative effects of IOp are statistically different from zero only in the first period immediately after it is observed, while there is not a statistically significant relationship between IOp and the annual growth rate in the last three periods (years 3 and 4). However, the coefficient of IOp may still be including some endogenous impact of residual inequality, making it biased. In fact, when accounting for the latter, although the same pattern emerges, the second local projection becomes statistically significant, as displayed in Figure 1(b). In particular, a one-point increase in the Gini index of IOp in the baseline period (year zero) implies a decrease of about 0.22 p.p. and 0.14 p.p. in the annual growth rate after one year and two years, respectively. In the third and fourth years after, instead, the effect goes statistically to zero.

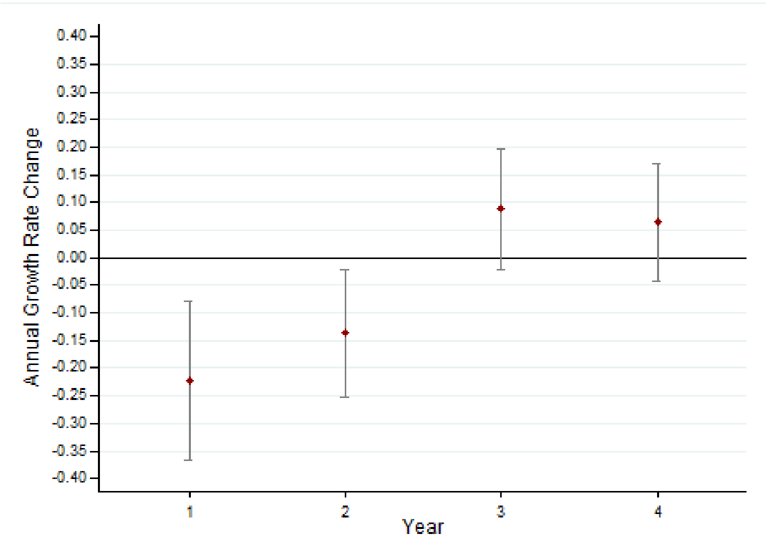
In Figure 2(a), instead, it is presented the evidence emerging from final specification including regressors selected by a Post-Double LASSO routine, estimated through OLS including only IOp. Again, a negative and statistically significant effect of inequality of opportunity on punctual growth rate emerges in the first two years after the former is observed, while it becomes statistically non-different from zero in the last two years. The inclusion of residual inequality in the model, as reported in Figure 2(b), does not changes this conclusion. More specifically, when total inequality is included in the regression, a one point increase in IOp Gini index in year zero implies about 0.18 p.p. decrease in the annual growth rate in the first year, and about 0.14 p.p. in the second year, while it fades to zero in the third and fourth years. Finally, mirroring the benchmark analysis, the model specification provided by the Post-Double LASSO routine is estimated in a RE framework, both excluding and including residual inequality as a covariate. The former case is presented in Figure 3(a). Also in this case, a negative and statistically significant effect arises, presenting a pattern that re-absorbs over time. The inclusion of residual inequality, showed in Figure 3(b), does not show any relevant change in the behaviour of IOp coefficients over time in respect of previous models, although the trend is less sloped in the first two years. In fact, a one-point increase in the IOp Gini index in the baseline period implies a decrease in the annual growth rate slightly higher than 0.15 p.p. in the subsequent year, and slightly less in the year after.

An overall conclusion arises. Isolating a circumstance-driven component within inequality of outcomes leads to a twofold finding. On one hand, it helps to clarify the link between the inequality and growth. On the other hand, digging further in the dynamic link between the two variables allows to disentangle the relationship over time, capturing the temporal pattern in which IOp affects growth. In particular, the negative effect, emerging from the first part of the analysis, is particularly relevant only in the first two years after inequality of opportunity is observed, while fading away as time passes. This evidence advocates for an immediate

Figure 1: *The Dynamic Effect of IOp on Growth (TWM Model)*

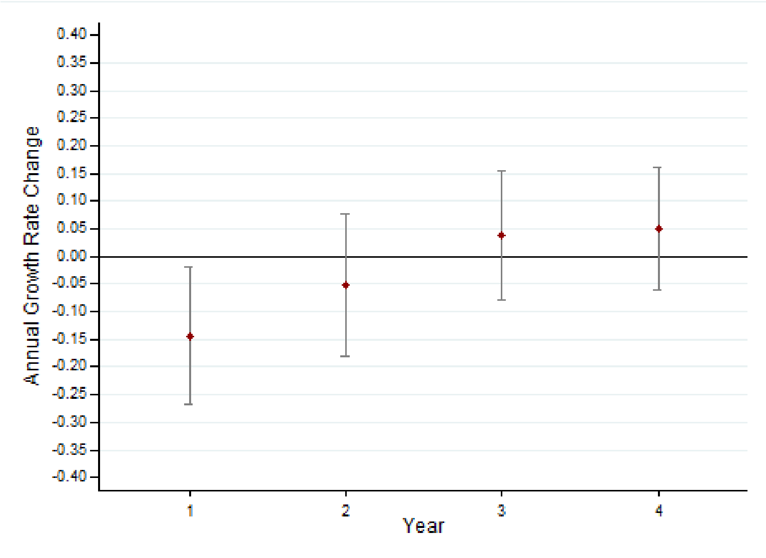


(a) Only IOp Included

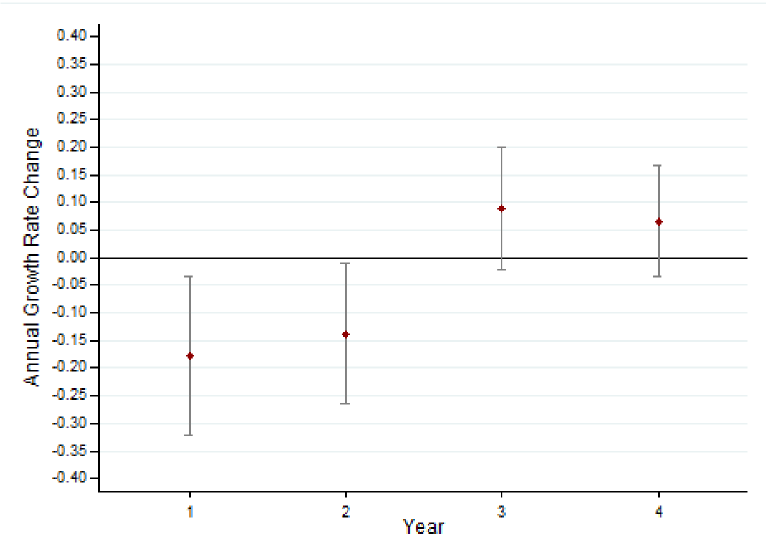


(b) Both IOp and TI included

Figure 2: *The Dynamic Effect of IOp on Growth (Post-Double LASSO Model - OLS)*

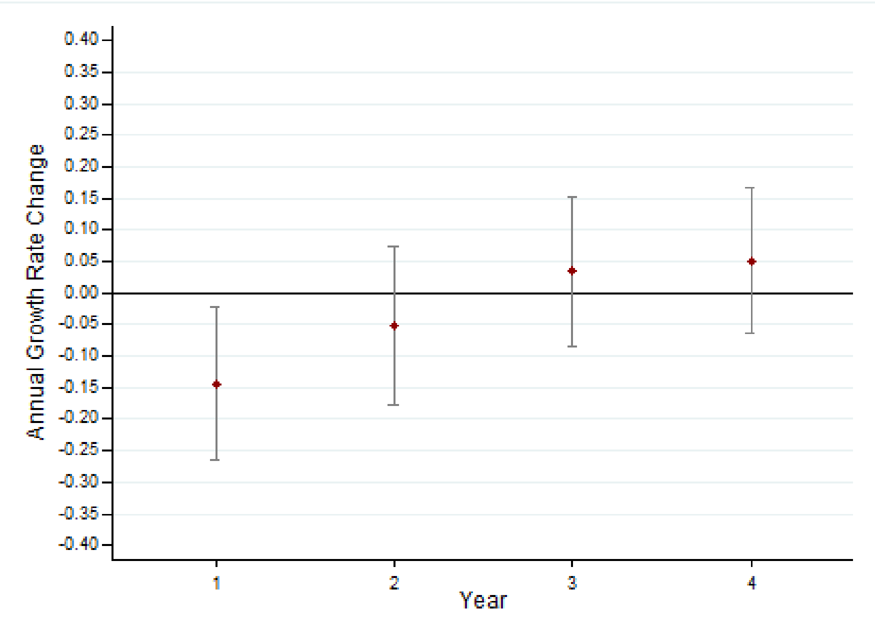


(a) Only IOp Included

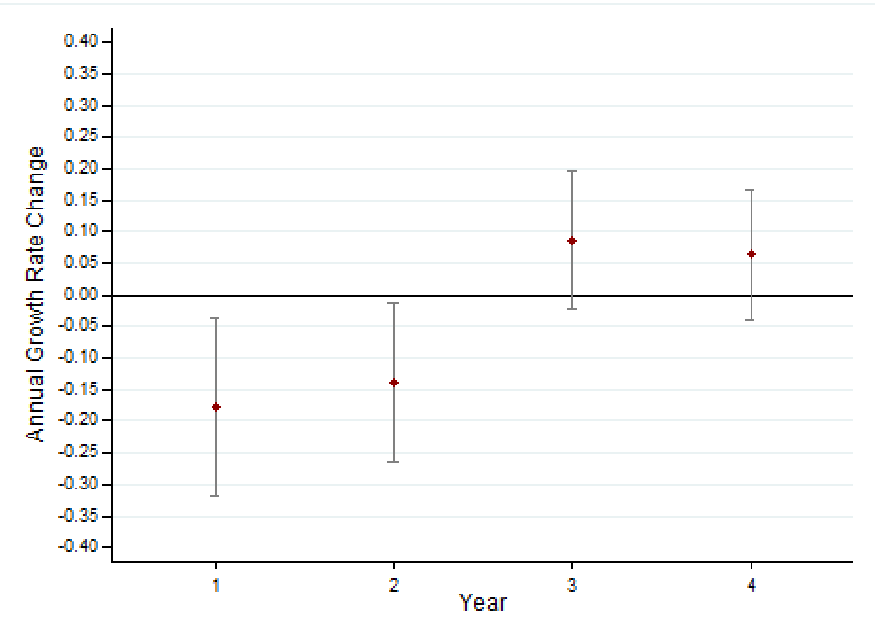


(b) Both IOp and TI included

Figure 3: *The Dynamic Effect of IOp on Growth (Post-Double LASSO Model - RE)*



(a) Only IOp Included



(b) Both IOp and TI included

relationship between inequality of opportunity and growth, emphasizing the role it may play, for instance, in (in)efficient allocation processes of resources, crucial engine for future growth.

7 Conclusions

The present work is positioned in the field of research investigating the relationship between inequality and economic growth. In particular, it attempts to explain the mixed findings in literature about the effect of inequality on economic performance, by arguing that such ambiguity arises from the adopted definition of inequality. More specifically, it supports the hypothesis that isolating a circumstance-driven (unfair) component of inequality of outcomes may help to disentangle the puzzle emerging from previous works. In this regard, inequality of opportunity is tested as a driver of economic growth, both singularly and jointly with total inequality of outcomes.

This article presents several novelties. First, machine learning algorithms – Post-Double LASSO – are implemented to select the optimal set of regressors while addressing endogeneity arising from omitted variable bias. This is particularly relevant in consideration of the many different model specifications adopted by previous works. Second, an original overview of the dynamic pattern of IOp effects on growth is proposed through local projections. Third, the estimates of inequality of opportunity used in the empirical analysis rely on a much broader set of circumstances in respect of previous published works. In particular, they are capture both the socio-economic and demographic backgrounds of individuals when they were in young age. This allows for a more accurate isolation of inequality of opportunity from total inequality. Fourth, the inequality decomposition is performed at the regional level. This allows to implement a novel investigation in regional Europe of the relationship between inequality of opportunity and growth, implying a higher degree of detail in considering territorial heterogeneities. Fifth, agnosticism about the relevant aspects to the growth engine is adopted, including all those fragmentarily proposed in previous studies.

The benchmark analysis provides strong evidence of a negative and statistically significant effect of inequality of opportunity on average economic growth at the regional level in Europe in the subsequent four years. Such results is robust regardless the application of machine learning regularization routines and the inclusion of residual inequality. It also emerges a dynamic effect of inequality of opportunity on economic performance. In particular, the growth is affected by IOp in the first two time horizons after the latter is observed, while the negative impact fades away in the third and fourth years. As in the previous case, this finding is robust to the implementation of machine learning routines and to the inclusion of residual inequality. Further research may require greater provision of retrospective data, both in terms of type of information and observation period span. This would allow to investigate the relationship in longer longitudinal datasets and in higher sample dimensions, while including more information about individuals' socio-economic background. Still, the emerging evidence is the best that can be achieved about the inequality-growth nexus, considering the constraint of current data availability.

References

Aiyar, S. and C. Ebeke (2020). 'Inequality of Opportunity, Inequality of Income and Economic Growth'. *World Development 136*.

Alesina, A., R. Di Tella, and R. MacCulloch (2004). 'Inequality and Happiness: are Europeans and Americans different?'. *Journal of Public Economics 88*, 2009–2042.

Alesina, A. and R. Perotti (1996). 'Income Distribution, Political Instability and Investment'. *European Economic Review 40*, 1203–1228.

Alesina, A. and D. Rodrik (1994). 'Distributive Politics and Economic Growth'. *The Quarterly Journal of Economics 109*, 465–490.

Anscombe, G. (1958). 'Modern Moral Philosophy'. *Philosophy 33(124)*, 1–19.

Banerjee, A. and E. Duflo (2003). 'Inequality and Growth: what can the data say?'. *Journal of Economic Growth 8*, 267–299.

Barro, R. (2000). 'Inequality and Growth in a Panel of Countries'. *Journal of Economic Growth 5*, 5–32.

Belloni, A., V. Chernozhukov, I. Fernandez, and C. Hansen (2017). 'Program evaluation with highdimensional data'. *Econometrica 85*, 233–298.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). 'Inference on treatment effects after selection amongst high-dimensional controls'. *Review of Economic Studies 81(2)*, 608–650.

Berg, A., J. Ostry, and C. Tsangarides (2014). 'Redistribution, Inequality, and Growth'. *IMF Staff Discussion Note*.

Blackburn, M. (2007). 'Estimating Wage Differentials Without Logarithms'. *Labour Economics 14(1)*, 73–98.

Bolton, G. and A. Ockenfels (2000). 'A Theory of Equity, Reciprocity and Competition'. *American Economic Review 90(1)*, 166–193.

Bradbury, K. and R. Triest (2016). 'Inequality of Opportunity and Aggregate Economic Performance'. *The Russel Sage Foundation Journal of the Social Sciences 2(2)*, 178–201.

Carranza, R. (2020). 'Inequality of Outcomes, Inequality of Opportunity, and Economic Growth'. *ECINEQ WP 534*.

Chen, S., V. Chernozhukov, and I. Fernandez-Val (2019). 'Mastering Panel Metrics: Causal Impact of Democracy on Growth'. *AEA Papers and Proceedings 109*, 77–82.

Deininger, K. and L. Squire (1996). 'A New Dataset Measuring Income Inequality'. *World Bank Economic Review 10*, 565–591.

Dworkin, R. (1981). 'What Is Equality? Part 2: Equality of Resources'. *Philosophy and Public Affairs 10(4)*, 283–345.

Ferreira, F. and J. Gignoux (2011). 'The Measurement Of Inequality Of Opportunity: Theory and an Application To Latin America'. *Review of Income and Wealth 57*, 622–657.

Ferreira, F., C. Lakner, M. Lugo, and B. Ozler (2017). 'Inequality of Opportunity and Economic Growth: how much can cross-country regressions really tell us?'. *Review of Income and Wealth 64*, 800–827.

Ferrer-I-Carbonell, A. and X. Ramos (2014). 'Inequality and Happiness'. *Journal of Economic Surveys 28(5)*, 1016–1027.

Filmer, D. and L. Pritchett (2001). 'Estimating Wealth Effects Withour Expenditure Data or Tears: an application to educational enrollments in states of India'. *Demography 38(1)*, 115–132.

Forbes, K. (2000). 'A Reassessment of the Relationship Between Inequality and Growth'. *American Economic Review 90*, 869–887.

Galor, O. and O. Moav (2004). 'From Physical to Human Capital Accumulation: inequality and the process of development'. *The Review of Economic Studies 71*, 1001–1026.

Galor, O. and J. Zeira (1993). 'Income Distribution and Macroeconomics'. *The Review of Economic Studies 60*, 35–52.

Kaldor, N. (1957). 'A Model of Economic Growth'. *The Economic Journal 67*, 591–627.

Keynes, J. (1936). 'The General Theory of Employment, Interest, and Money'. *UTET 2012*.

Lewis, P. (1954). 'Economic Development with Unlimited Supplies of Labour'. *The Manchester School 22(2)*, 139–191.

Li, H. and H. Zou (1998). 'Income Inequality is not Harmful for Growth: theory and evidence'. *Review of Development Economics 2*, 318–334.

Marrero, G. and J. Rodriguez (2013). 'Inequality of Opportunity and Growth'. *Journal of Development Economics 104*, 107–122.

Marrero, G., J. Rodriguez, and V. R. (2016). 'Unequal Opportunity Unequal Growth'. *World Bank Policy Research Working Paper 7853*.

Mundlak, Y. (1978). 'On the Pooling of Time Series and Cross-Section Data.'. *Econometrica 46(1)*, p.69–85.

Murphy, K., A. Shleifer, and R. Vishny (1989). 'Income Distribution, Market Size and Growth'. *The Quarterly Journal of Economics 104*, p.537–564.

Niehues, J. and A. Peichl (2014). 'Upper Bounds of Inequality of Opportunity: Theory and Evidence for Germany and US'. *Social Choice and Welfare 43(1)*, 73–99.

Pasinetti, L. (1962). 'Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth'. *Review of Economic Studies 29*, 267–279.

Perotti, R. (1993). 'Political Equilibrium, Income Distribution, and Growth'. *The Review of Economic Studies 60*, 755–776.

Perotti, R. (1996). 'Growth, Income Distribution and Democracy'. *Journal of Economic Growth 1*, 149–187.

Persson, T. and G. Tabellini (1994). 'Is inequality harmful for growth?'. *American Economic Review 84*, 601–624.

Roemer, J. (1998). 'Equality of Opportunity'. *Harvard University Press*.

Saint-Paul, G. and T. Verdier (1993). 'Education Growth and Democracy'. *Journal of Development Economics 42*, 399–407.

Szèkeli, M. (2003). 'The 1990s in Latin America: another decade of persistent inequality, but with somewhat lower poverty'. *Journal of Applied Economics 6*, 317–339.

Teyssier, G. (2013). 'Inequality, Inequality of Opportunity, and Growth: what are we talking about? Theory and empirical investigation in Brazil'. *Université Paris I Pantheòn-Sorbonne Unpublished Manuscript.*

Urminky, O., V. Chernozhukov, and C. Hansen (2016). 'Using Double-LASSO Regression for Principled Variable Selection'. *Available at SSRN: Urminsky, Oleg and Hansen, Christian and Chernozhukov, Victor, Using Double-Lasso Regression for Principled Variable Selection (2016). Available at SSRN: https://ssrn.com/abstract=2733374 or http://dx.doi.org/10.2139/ssrn.2733374*.

Voitchovsky, K. (2012). 'The Oxford Handbook of Economic Inequality'. *Oxford University Press*.

Wooldridge, J. (2021). 'Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators.'. *Available at SSRN 3906345*.