APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2024.2424948

New estimates for inequality of opportunity in Europe using elastic net algorithms

Sara Ayllón [a], Pablo Brugarolas [b] and Samuel Lado [b]

[a]Department of Economics, Universitat de Girona, IZA and EQUALITAS, Girona, Spain; [b]Department of Economics, Universitat de Girona and EQUALITAS, Girona, Spain

ABSTRACT
This paper provides new estimates for inequality of opportunity (IOp) in Europe between 2005 and 2019, using data from EU-SILC and elastic net algorithms. We document three different trends: some countries showed significant improvement over time; others saw a notable increase in IOp; and yet others reversed the rise in inequality post-Great Recession. Importantly, our new machine-learning estimates show consistency with more established approaches.

KEYWORDS
Inequality of opportunity; elastic net algorithms; Europe; EU-SILC

JEL CLASSIFICATION
D63; D39

I. Introduction

The economic literature on inequality of opportunity (IOp) differentiates between 'circumstances' – factors beyond one's control, like race or parental education – and 'efforts', such as career choice or education, over which an individual has (greater) control (Roemer 1998). Inequality of opportunity is measured solely by the impact of circumstances.

Early studies, following Bourguignon, Ferreira, and Menéndez (2007), estimated income IOp using ordinary least squares (OLS) regression, predicting income from observed circumstances. However, this method, constrained as it is by unobserved factors, often underestimates IOp, providing only lower bound estimates (Balcázar 2015). Moreover, reliance on OLS has been criticized for upward bias due to overfitting, as too many circumstances are often included (Brunori, Peragine, and Serlenga 2019).

To address upward bias, recent advances have employed machine-learning (ML) techniques to estimate IOp. Some studies have implemented the latent class approach, where groups of individuals exposed to the same circumstances ('types') are obtained by estimating an optimal number of latent types that is a function of the set of observed circumstances (Carrieri, Davillas, and Jones 2020; Brunori, Trannoy, and Guidi 2021; Li Donni, Rodríguez, and Rosa Dias 2015). This approach is, however, prone to underfitting the data, leading to downward-biased estimates of IOp (Brunori, Hufe, and Mahler 2023). Another strand of literature has relied on tree-based methods, which partition the data into non-overlapping regions on the basis of circumstance variables and make predictions based on these groupings, effectively dealing with upward bias (Brunori, Hufe, and Mahler 2018, 2023; Brunori and Neidhöfer 2021; Carranza 2023).

Despite progress, Brunori, Hufe, and Mahler (2023) recently pointed out that ML-derived IOp estimates lack internal consistency and correlate only moderately with traditional social-mobility measures, like intergenerational income elasticity (IGE). We contribute to the literature by proposing the use of a novel algorithm – the elastic net – which belongs to the promising (though less explored) approach of regularization techniques (see Hufe, Peichl, and Weishaar 2022, for an exception). Our elastic net IOp estimates address the methodological discrepancies in the measurement of the lower bound in IOp, providing remarkable internal consistency with more established approaches and the IGE literature.

II. Method

Following standard procedures, we adopt an ex-ante definition of IOp. According to this approach,

CONTACT Sara Ayllón sara.ayllon@udg.edu Department of Economics, Universitat de Girona, IZA and EQUALITAS, Girona 10, 17003, Spain

a population of individuals i within the set $\{1, \ldots, N\}$ with income levels $\{y_1, \ldots, y_n\}$ is classified into distinct types $\{T_1, \ldots, T_k\}$ that share the same circumstances. The presence of IOp is determined by the differences observed in average income across the different types. In particular, a counterfactual distribution is estimated by replacing individual incomes with the mean income of the corresponding type: $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_n\}$. Consequently, IOp can be evaluated as the inequality of $\tilde{\mu}_i$, $IOp = I(\tilde{\mu}_1, \ldots, \tilde{\mu}_n)$, through an inequality index $I(\cdot)$. Along with much of the recent literature, we employ the Gini coefficient as the chosen inequality index (e.g. Brunori, Hufe, and Mahler 2023; Carranza 2023).

Various machine-learning approaches have been advanced in the literature to partition the population into the types required for computation of the counterfactual distribution $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_n\}$ while avoiding the upward-bias problem. In this paper, we propose to use the elastic net to calculate lower bound estimates for IOp. Elastic net estimations, developed by Zou and Hastie (2005), incorporate a mixture of lasso and ridge penalties. This flexibility ensures that the model benefits from both regularization techniques, adapting to different data structures and correlation patterns. As a result, generalization is improved and out-of-sample predictions are more accurate, ultimately enhancing estimation of the lower bound. We first run the following model:

$$\hat{\upbeta} \equiv \underset{\beta}{argmin} \left(||y - X\upbeta||^2 + \lambda_2 ||\upbeta||^2 + \lambda_1 ||\upbeta||_1 \right) \quad (1)$$

where $\hat{\beta}$ is the vector of estimated coefficients that minimizes the objective function, y represents the outcome variable, X is the set of circumstances that explain the outcome, and $||y - X\upbeta||^2$ is the OLS loss function that measures the sum of squared errors between the observed and predicted outcomes. $\lambda_2||\upbeta||^2$ and $\lambda_1||\upbeta||_1$ introduce penalties: the former penalizes the magnitude of the coefficients (ridge penalty), while the latter penalizes the absolute values of the coefficients (lasso penalty), driving some coefficients to zero. λ_1 and λ_2 control the trade-off between the two regularization techniques, and they are chosen through 5-fold cross-validation. We then use the resulting vector $\beta_j^{elasticnet}$ to construct the counterfactual distribution $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_n\}$ from which the measure of IOp will be derived[1]

$$\widetilde{\mu_{it}^{LB}} = \exp\left\{ \alpha^{\widehat{\text{elastic net}}} + \upbeta^{\widehat{\text{elastic net}}} * Ci + \frac{\sigma^2}{2} \right\} \quad (2)$$

where the counterfactual distribution is obtained from combining the predicted values from the elastic net model ($\alpha^{\widehat{\text{elastic net}}}, \upbeta^{\widehat{\text{elastic net}}}$) with the set of individual circumstances, Ci. This construction reflects the income distribution that would exist if all individuals shared the same circumstances.

III. Data

To measure the effect of circumstances on income, we use data from the European Union – Statistics on Income and Living Conditions (EU-SILC). Our outcome of interest is household equivalized disposable income, using the modified OECD equivalence scale. This is the preferred outcome in most literature on the measurement of IOp. Our sample consists of individuals aged 25–55 with positive income.

Lower bound estimates can be derived from the cross-sectional component of the database for 2005, 2011 and 2019, as these waves include ad-hoc modules with retrospective information on the respondent's parental background and the household in which she grew up. To make the results comparable across waves, we use the same set of circumstances – detailed in Table S1 in the online Appendix.

IV. Results

Figure 1 presents the lower bound IOp estimates in various European countries for the years 2005, 2011 and 2019, calculated using the elastic net method.[2] Consistent with other research, we find

[1]We estimate a second lower bound implementing a post-OLS elastic net, retaining the subset of circumstances whose coefficients did not shrink to zero in the elastic net estimation. The results are similar to those below and are available upon request.:

[2]The figure includes bootstrapped standard errors based on 200 re-samples of the data.

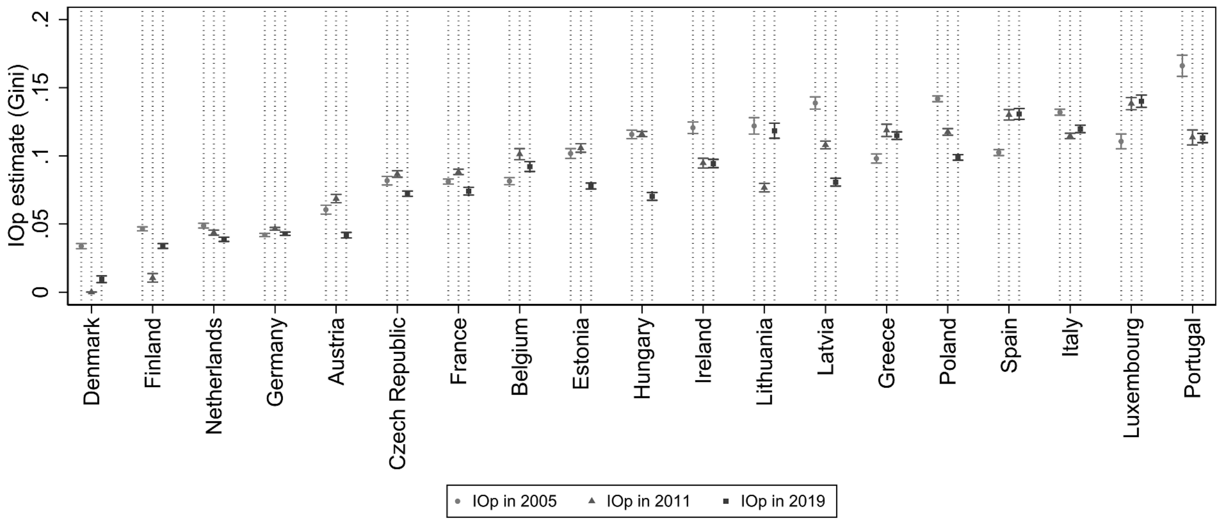

Figure 1. Lower bound IOp estimates using elastic net, Europe, 2005, 2011 and 2019.
Note: Vertical lines represent confidence intervals at 95%. Countries are sorted based on their mean value of IOp. Source: Authors' computation, using data from the EU-SILC.

that Northern and Central European countries – such as Denmark, Finland, the Netherlands, Germany and Austria – exhibit lower levels of IOp. By contrast, higher levels are evident in a group comprising Mediterranean and Eastern European countries – including Spain, Italy, Lithuania, Greece, Portugal and Poland.[3] It is immediately clear that a large number of countries show a persistent level of IOp over the period examined, as indicated by the minimal spread of their estimates over time. There are, however, few exceptions with Portugal and Latvia signalling a considerable improvement. The trajectories of Luxembourg and Spain are particularly negative over the period: they moved from a mid-tier position in 2005 to top of the IOp rankings by 2019. Other countries saw their IOp rise during the Great Recession, but then improve by decade's end: the

Table 1. Rank correlations (Spearman's *r*) of ML-based approaches.

	Existing studies			This paper
	CV_b	FOR_c	FOR_b	EN
A-Equality of opportunity in 2011 (MLD)				
EN	–			1.000
CV_b	1.00		–	0.88
FOR_c	0.58	1.00	–	0.47
B-Equality of opportunity in 2011 (Gini)				
FOR_b	–	–	1.00	0.86
C-Equality of opportunity in 2019 (MLD)				
FOR_c	–	1.00	–	0.63
D-Intergenerational income elasticity (9 countries)				
IGE	–	–	0.72	0.88

Panels (A) to (C) are based on the intersection of countries included in this paper across studies that provide MLD-based IOp estimates (panels (A) and (C)) or Gini-based IOp estimates (panel (B)). Ranks in Stuhler (2018) and Carmichael et al. (2020) are calculated from consensus estimates of IGE.
EN = Elastic net, CV_b = Cross-validation in Brunori, Peragine, and Serlenga (2019), FOR_c = Forest in Carranza (2023) and FOR_b = Forest in Brunori, Hufe, and Mahler (2023).

[3]We conducted a robustness test to evaluate whether missing values across variables could influence the estimates. We recalibrated the IOp levels across countries, imposing a maximum threshold of 20% for missing values per variable. The findings are reassuringly consistent, indicating that this phenomenon does not compromise our results. Details are available upon request.

trend for Hungary is especially pronounced. Overall, the analysis reveals a diverse range of trends in the IOp progression across different countries.

Furthermore, Brunori, Hufe, and Mahler (2023) show, through rank correlations, that the existing literature on IOp in Europe is not internally consistent and is only moderately consistent with IGE rankings (their random forest approach being a notable exception). We update this exercise to test how reliable the novel elastic net method is. Table 1 shows that our elastic net estimates correlate closely with the forest estimates of Brunori, Hufe, and Mahler (2023), the cross-validation (CV) approach (Brunori, Peragine, and Serlenga 2019) and IGE rankings (Carmichael et al. 2020; Stuhler 2018). Conversely, the random forest estimates of Carranza (2023) appear to be more inconsistent with both elastic net and CV. The proposed new algorithm for measuring IOp demonstrates a remarkable improvement in the consistency of employing ML to estimate unfair inequality.

Source: Authors' computation and estimates from the studies included.[4]

V. Conclusions

This paper provides new estimates for IOp in Europe between 2005 and 2019. We employ a machine-learning technique – elastic net – which offers a greater degree of flexibility than methods used previously, leading to improved generalization. We document three divergent evolutionary paths, with countries like Portugal showing significant improvement over time; others, like Luxembourg or Spain, experiencing a notable increase in IOp; and a third group, including Hungary, reversing the rise in inequality post-Great Recession.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Support from the Instituto de Estudios Fiscales and the projects 2021SGR570 and PID2022-137352NB-C41 [funded by MCIN/AEI/10.13039/501100011033] is acknowledged. Brugarolas acknowledges support from IFUdG2022 and Lado from FPU21/00242.

ORCID

Sara Ayllón http://orcid.org/0000-0002-3338-1183
Pablo Brugarolas http://orcid.org/0000-0001-5290-4755
Samuel Lado http://orcid.org/0000-0002-7290-7598

References

Balcázar, C. F. 2015. "Lower Bounds of Inequality of Opportunity and Measurement Error." *Economics Letters* 137:102–105. https://doi.org/10.1016/j.econlet.2015.10.026.

Bourguignon, F., F. H. G. Ferreira, and M. Menéndez. 2007. "Inequality of Opportunity in Brazil." *Review of Income Wealth* 53 (4): 585–618. https://doi.org/10.1111/j.1475-4991.2007.00247.x.

Brunori, P., P. Hufe, and D. Mahler. 2018. "The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees." *World Bank Policy Research Working Paper*, 8349.

Brunori, P., P. Hufe, and D. Mahler. 2023. "The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees and Forests." *The Scandinavian Journal of Economics* 125 (4): 900–932. https://doi.org/10.1111/sjoe.12530.

Brunori, P., and G. Neidhöfer. 2021. "The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach." *Review of Income and Wealth* 67 (4): 900–927. https://doi.org/10.1111/roiw.12502.

Brunori, P., V. Peragine, and L. Serlenga. 2019. "Upward and Downward Bias When Measuring Inequality of Opportunity." *Social Choice and Welfare* 52 (4): 635–661. https://doi.org/10.1007/s00355-018-1165-x.

Brunori, P., A. Trannoy, and C. F. Guidi. 2021. "Ranking Populations in Terms of Inequality of Health Opportunity: A Flexible Latent Type Approach." *Health Economics* 30 (2): 358–383. https://doi.org/10.1002/hec.4185.

Carmichael, F., C. K. Darko, M. G. Ercolani, C. Ozgen, and W. S. Siebert. 2020. "Evidence on Intergenerational Income Transmission Using Complete Dutch Population Data." *Economics Letters* 189:108996. https://doi.org/10.1016/j.econlet.2020.108996.

Carranza, R. 2023. "Upper and Lower Bound Estimates of Inequality of Opportunity: A Cross-National Comparison for Europe." *Review of Income and Wealth* 69 (4): 838–860. https://doi.org/10.1111/roiw.12622.

Carrieri, V., A. Davillas, and A. M. Jones. 2020. "A Latent Class Approach to Inequity in Health Using Biomarker

[4]We thank Paolo Brunori, Paul Hufe and Daniel Mahler for sharing the IGE data with us.

Data." *Health Economics* 29 (7): 808–826. https://doi.org/10.1002/hec.4022.

Hufe, P., A. Peichl, and D. Weishaar. 2022. "Lower and Upper Bound Estimates of Inequality of Opportunity for Emerging Economies." *Social Choice and Welfare* 58 (3): 395–427. https://doi.org/10.1007/s00355-021-01362-7.

Li Donni, P., J. G. Rodríguez, and P. Rosa Dias. 2015. "Empirical Definition of Social Types in the Analysis of Inequality of Opportunity: A Latent Classes Approach." *Social Choice and Welfare* 44 (3): 673–701. https://doi.org/10.1007/s00355-014-0851-6.

Roemer, J. E. 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Stuhler, J. 2018. *A Review of Intergenerational Mobility and Its Drivers*. Luxembourg: Publications Office of the European Union.

Zou, H., and T. Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society Series B, Statistical Methodology* 67 (2): 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.