PAPELES DE TRABAJO

6/2023

The level of inequality of opportunity in Spain: an estimation using Artificial Intelligence

SARA AYLLÓN (*)

Department of Economics, IZA and EQUALITAS. Universitat de Girona

PABLO BRUGAROLAS (**)

Department of Economics. Universitat de Girona

SAMUEL LADO (***)

Department of Economics. Universitat de Girona

(*) Sara Ayllón. Support from the Instituto de Estudios Fiscales, REQ 2021 (*'Plan de Recuperación, Transformación y Resiliencia'*, Ministerio de Universidades and NextGenerationEU) and the projects PID2019-104619RB-C43 (funded by MCIN/AEI/10.13039/501100011033) and 2021SGR570 is acknowledged. Any errors or misinterpretations are our own.

(**) Pablo Brugarolas acknowledges support from IFUdG 2022 ('Ayudas para la contratación de personal investigador en formación de la Universitat de Girona, Universitat de Girona) and the project 2021SGR570.

(***) Samuel Lado acknowledges support from FPU 2021 ('Ayudas para la Formación de Profesorado Universitario', Ministerio de Universidades) and the project 2021SGR570.

Avda. Cardenal Herrera Oria, 378, 28035 Madrid

I.S.S.N.: 1578-0252

N.I.P.O.: 140-23-017-0

ÍNDICE

1. INTRODUCTION
2. LITERATURE REVIEW
   2.1. Recent advances in the measurement of the lower bound of inequality of opportunity: machine learning techniques
   2.2. Recent case studies in the measurement of the upper bound of IOp
3. OUR PROPOSAL
   3.1. Elastic net and post-OLS elastic net estimations for the lower bound of inequality of opportunity
   3.2. Effort and the measurement of the upper bound of inequality of opportunity
4. DATA
5. RESULTS
   5.1. Lower bound
   5.2. Upper bound
6. CONCLUSIONS

References

APPENDIX A

APPENDIX B

1. INTRODUCTION

The principle of equality of opportunity aims to ensure that adults who exert the same level of effort are capable of achieving the same outcomes, regardless of the circumstances in which they were raised. Accomplishing such a goal is regarded as a success in any society because it means that the income level that a person obtains as an adult is not conditioned by the resources of the environment in which s/he was born. This is one of the objectives of inclusion and redistribution policies such as early childhood education, scholarship programmes, or the recently developed Minimum Living Income in Spain. With measures like these, Spanish society seeks to guarantee that all households, especially those with children, can ensure that their achievements depend to a great extent on their efforts rather than the circumstances they inherit.

One of the main reasons Spain is not particularly effective in reducing inequality of opportunity (IOp) is that the indicators developed by the academic literature to document this issue have not gained substantial traction in public debate. Political parties, the media, social actors, and citizens often rely on the Consumer Price Index (IPC), Gross Domestic Product (GDP), or the Gini index to support their arguments. However, it is anomalous in our society to resort to indicators of inequality of opportunities to enrich these debates. This can be partly explained by their relatively recent development: the metrics that empirical evidence handles to measure the level of inequality of opportunity in our societies have been somewhat less consolidated over these years than those existing to measure other phenomena such as unemployment, economic growth, or price levels.

In the last decade, however, there have been several methodological contributions to measuring inequality of opportunities that position their metrics in a more consolidated and mature state – as we will further discuss in the following sections. The main objective of this research is to bring together these contributions for the first time to offer the most rigorous possible measurement of IOp in Spain, and in comparison with the rest of the European countries. We contribute to the literature in the following ways. First, we aim to address the methodological discrepancies in the measurement of the lower bound in IOp observed in recent literature (Brunori *et al*., 2023). To this end, we propose the use of a novel machine learning algorithm, the elastic net. This method demonstrates strong consistency with other approaches, such as conditional inference regression forests, and it is also aligned with intergenerational income elasticity (IGE) rankings. Second, we aim to improve the measurement of the upper bound estimate of IOp by carefully considering multiple variables that can potentially measure effort (and the indirect effect of circumstances on effort) and, as a result, discern the importance of effort on the measurement of IOp. As a result, this paper provides new lower and upper bound estimates of IOp in Spain and across several European countries for 2005, 2011 and 2019.

Our main findings regarding the lower bound are threefold. First, we document a substantial rise in unfair inequalities contributing to income inequality in Spain, with the share of total inequality attributed to IOp increasing from 11.5% in 2005 to 17.6% in 2019. Second, we show that Spain has seen its relative position worsen vis-à-vis other European countries, moving from the sixth to the fourth European country with a higher IOp. Third, and on the methodological side, we find that our elastic net estimates are remarkably aligned with those derived from other methods in the existing literature. Regarding the upper bound, we document that 68% of overall inequality in Spain is attributed to IOp, and this proportion has been increasing since 2005. When considering Europe as a whole, Spain held the eleventh position in terms of relative IOp in 2019. From a methodological perspective, we confirm previous findings in the literature: the consideration of the indirect effects of circumstances on effort or ignoring them yields surprisingly similar results.

Following this introduction, Section 2 of this paper provides a review of the economic literature on IOp with particular emphasis on the recent advances in the measurement of IOp using machine learning techniques and on the most recent case studies developed by other researchers. Section 3 provides all the details regarding the new approaches we propose to measure IOp. Section 4 gives an account of the dataset used and summary statistics. Section 5 presents the main findings, while Section 6 concludes and provides policy recommendations.

2. LITERATURE REVIEW

The economic literature on IOp argues that people's outcomes, such as their income or health, are determined by two sets of factors, 'circumstances' and 'efforts' (Fleurbaey, 1994; Roemer, 1993, 1998). Circumstances are factors beyond a person's control, such as race, socioeconomic and cultural status, and level of cultural development. Conversely, efforts would be the variables that the person has (greater) control over, such as her career of choice, the number of hours worked, or her investment in human capital. Consequently, total inequality is the end product of combining IOp and inequality of effort. IOp arises only when inequality is due to differences in circumstances beyond individual control.

Following the seminal contribution of Bourguignon *et al.* (2007), an early empirical literature provided the first estimates of income IOp – see Ferreira and Gignoux (2011) for an in-depth discussion. This approach quantifies the variability due to the vector of observed circumstances by ordinary least squares regression (OLS). As an illustration, consider the study by Marrero and Rodríguez (2012), which estimates IOp in income across European countries based on five circumstances experienced during childhood: the father's educational level, the mother's educational level, the father's type of occupation, immigrant background and the household's financial situation. Given that many relevant circumstances for predicting the outcome of interest are not observed, estimates based on this approach are typically considered to underestimate the actual level of IOp in a given society and thus are taken as lower bound estimates (Balcázar, 2015; Hufe *et al*., 2017).

To derive a measure of IOp, studies following this approach run an OLS regression that predicts income based on these factors beyond individual responsibility. Subsequently, they apply a measure of inequality to the estimated prediction, typically the Gini coefficient or the mean log deviation (MLD) and take the resulting counterfactual income distribution as the estimate of absolute IOp in that society. There will be IOp if the absolute IOp index is greater than 0, indicating that people with different inherited circumstances do not have the same chances of achieving the same income level.[1] These early empirical contributions tend to place Spain as one of the countries in Europe with the highest level of IOp. Marrero and Rodríguez (2012) estimate that Spain had, in 2005, the third highest level of total IOp in Europe, surpassed only by Lithuania and Portugal. The authors show how absolute inequality in Spain that year was well above that of other countries as France, Germany, and Italy. In relative terms, they estimate that Spain was the fourth most unequal European country in terms of IOp. For 2011, Palomino *et al*. (2019)'s analysis maintains the same five circumstances as the previous study, adding gender as an additional sixth circumstance. They show that Spain remained in 2011 as one of the countries with the highest level of IOp in Europe, ranking fifth. At first glance, comparing their relative estimate to the 2005 results reveals some improvement. In 2011, Spain is ranked tenth in Europe for having the highest inequality attributed to differences in circumstances of origin. However, the authors stress that this improvement is not due to a decrease in IOp (which remains constant with respect to 2005) but to an increase in overall inequality in Spain in these years due to the 2008 economic crisis. This increase in IOp has led to a decrease in the relative IOp ratio (IOR).[2]

However, some recent methodological contributions to the measurement of IOp have shown that this early OLS-based approach faces both downward and upward biases. Brunori *et al*. (2019) discuss the trade-off between the downward bias due to partial observability of circumstances, as introduced above and the upward bias, which contaminates estimates of IOp through raising sampling variance. Their study illustrates that as the list of observed circumstances becomes more extensive, the explanatory power of the model measuring IOp artificially increases, even if the incorporated characteristics are irrelevant. In other words, a researcher adopting this procedure is at risk of overestimating the role of circumstances in explaining inequality in the outcome of interest, such as income.

Furthermore, about a decade ago, Niehues and Peichl (2014) acknowledged that up to that point, all estimates of IOp were lower bounds and proposed the estimation of an upper bound which would take into account the maximum value of (observed and unobserved) circumstances – which they assumed to be not only exogenous but time-constant. Using longitudinal data, their method first estimates a time-invariant individual fixed effects (FE) which, by definition, comprises all time-invariant variables; and, in a second step, such effect is used to compute the maximum extent of inequality that can be attributed to IOp (as similarly done when computing the lower

[1] Most studies also use a second concept, described as relative IOp or IOp ratio (IOR). For a given country, this is simply the ratio of IOp over its level of total inequality. It thus estimates the proportion of total inequality that can be attributed to IOp.

[2] A recent study by FILAURO *et al*. (2023) also estimates IOp, providing both country-specific estimates and a novel pan-European analysis. Notably, the European Union is treated as a single entity, and the researchers incorporate the country of birth within the set of circumstances that are beyond individual control. Leveraging data from the EU-SILC and employing a multilevel model, they show that pan-European inequality of opportunity in individual earnings is much higher than any other country-specific estimates (in fact, pan-European inequality of opportunity is higher than the average inequality of opportunity between EU countries). Nonetheless, their findings reveal a declining trend, emphasising a process of convergence within Europe. Regarding the country-specific estimates, the results align with previous studies, indicating that, on average, about 40 percent of earnings inequality can be attributed to individual circumstances, with variations observed among EU countries.

bound). However, an important drawback of the method is that the individual time-constant effect not only includes time-invariant circumstance variables but also effort variables that do not change over time.[3] In this respect, the authors argue that the upward bias caused by the inclusion of time-constant effort may be partially compensated by the potential downward bias caused by existing time-varying circumstances.

Additionally, Niehues and Peichl (2014) propose the estimation of two upper bounds: one that fully compensates for the indirect effect of circumstances through effort on income and another that offers no compensation at all. In the first scenario, such indirect effects are treated as circumstances, while in the second scenario they are treated as effort. Their application to data from the Cross-National Equivalent File (CNEF) for Germany and the US allowed to understand the extent to which previous estimates in the literature were lower bounds of IOp and showed how fully compensating for the indirect effect of circumstances through effort on income increases more the estimate for IOp than when no compensation is accounted for.

In what follows, we summarise the recent advances in the measurement of the lower bound of inequality of opportunity using machine learning techniques. An additional section also provides details of recent studies that have estimated the upper bound of IOp in different contexts.

2.1. Recent advances in the measurement of the lower bound of inequality of opportunity: machine learning techniques

Brunori *et al*. (2019) pioneered the measurement of the lower bound of IOp through machine learning techniques, specifying various models ranging from simplistic predictions based on circumstances without interactions to complex models with all circumstances interacting. They selected the best model using cross-validation (CV), a statistical method that balances variance-bias by minimising the out-of-sample mean-squared error (MSE). The process involves splitting the sample into *folds* or subgroups and testing the model on each fold. The tuning parameter, the number of interactions included, is set, and the model is trained using all folds except one. The left-out fold contains the out-of-sample observations, and predictions generated from the remaining folds are compared to the true outcomes, yielding the out-of-sample MSE. This is repeated for all folds, and the overall MSE is calculated by averaging all the out-of-sample MSEs. The process is repeated with different levels of the tuning parameter, thus enabling the selection of a final income prediction model with the lowest cross-validated MSE.

Brzezinski (2020) used a different maximum likelihood (ML) algorithm to predict IOp in Europe, known as conditional inference regression trees (Hothorn *et al*., 2006). Trees predict the dependent variable income based on observable circumstances (or *features*, following the terminology in this field). To yield such prediction, the sample is divided into non-overlapping subgroups based on a partition of the circumstances' space. The prediction of each observation is just the mean

[3] Indeed, the authors acknowledge that the effect of circumstances on the outcome (e.g., income) could vary over time because of changes in the cultural norms or the institutions; and, that individuals' circumstances can also change because of, for example, the macroeconomic conditions.

value of the outcome variable in that group. Conditional regression trees condition each split on a sequence of statistical tests. First, one estimates the level of correlation between each circumstance and the outcome and selects the most correlated circumstance. The researcher then splits the sample into two groups according to the value of that first circumstance that yields the most significant differences in the expected outcome. For example, for the educational level of the father during childhood (coded as low, medium, or high), the routine will choose the partition that yields the most significant differences in expected income as an adult, say low vs medium/high. This procedure is done through the remaining circumstances, growing the tree until the researcher cannot reject the null that an additional split will yield a significant difference in the expected outcome.

In a recent contribution, Brunori *et al*. (2023) highlight that trees face several problems that make them a potentially weak ML algorithm, which might yield unstable predictions of income based on observed circumstances. Their paper proposes to estimate IOp in 31 European countries based on a conditional inference regression forest, which creates many trees and averages over all of these when making predictions. Forests are grown following the same procedure outlined above for trees, with two additional tweaks. First, each tree is estimated using a bootstrapped random subsample. Second, a random subset of circumstances is considered at each splitting point. Together these two tweaks address the drawbacks of conditional inference trees. First, the variance in each prediction's estimate is reduced by averaging over trees. Second, using subsets of circumstance variables enhances the possibility that any observation with informative content will eventually be recognised as a splitting variable. Carranza (2023) also employs a conditional inference regression forest to provide lower bound IOp estimates for Europe in 2005, 2011 and 2019.

Finally, Hufe *et al*. (2022) introduced a cross-validated lasso procedure to calculate lower bound measures of IOp for emerging economies. This procedure augments the OLS approach by adding a penalty term that navigates the trade-off between variance and bias. The lasso, or 'least absolute shrinkage and selection 'operator', is an estimation procedure that shrinks the coefficient estimates of less important variables towards zero, effectively removing them from the model. In the context of IOp, this can be useful in identifying and focusing on the most relevant predictors of income inequality. The optimal penalty term for the lasso procedure is determined through cross-validation, similar to previously discussed methods. In their application, Hufe *et al*. (2022) use this algorithm to identify and select the circumstances that are most significantly associated with income inequality. The selected circumstances are then used to construct a counterfactual income distribution, providing an estimate of the lower bound of IOp.

2.2. Recent case studies in the measurement of the upper bound of IOp

The recent literature that estimates the upper bound of IOp has mainly focused on the application to other countries of the method proposed by Niehues and Peichl (2014) in their pioneer work and much less on suggesting new estimation methods. Flatscher (2020) investigates the level of IOp in the United Kingdom, providing measurements for both the lower and upper bounds using the UK Household Longitudinal Study, a panel dataset spanning seven years. As Niehues and Peichl (2014), he estimates IOp using a FE model to account for unobserved circumstances that are time-invariant. However, he does not compensate for the indirect effects of circumstances on effort. Instead, he employs two approaches. First, he estimates a FE model while controlling for various effort variables, such as marital status, weekly working hours and region of residence. Second, he employs a FE model that does not explicitly account for effort, acknowledging that individuals may not have complete control over certain effort variables. His results show that the share of inequality due to circumstances ranges from 10.6 (lower bound) to 65.9 (upper bound) when using the gross income distribution. Based on Shapley's decomposition estimates, he concludes that gender and parental education play a significant role in driving IOp in the UK.

Hufe *et al*. (2022) provide new upper bound estimates of IOp for 12 emerging economies with data from the 90s until 2016, depending on the context and data availability. The authors conclude that downward bias dominates in these contexts, provided that the scope for overfitting circumstance information is less likely than in more developed economies, given the limitation of available variables. The sparsity of observable information also helps to explain the large differences between the lower and the upper bound estimates which, at the same time, are comparable to more developed economies.

Carranza (2023) is the first to offer an upper bound estimate for Spain. Using data from the EU-SILC, he provides lower and upper bound estimates of IOp for 32 European countries. As for the latter, he estimates the individual FE using a 3-year window.[4] Considering four measures of time-invariant effort, he provides three upper bound estimates of IOp. First, he accounts for both the direct and indirect influence of circumstances. Second, he completely removes the influence of efforts from his measure of circumstances. Third, he considers the indirect influence of circumstances on effort as a circumstance itself. His upper bound results show that IOp accounts from 20 percent to nearly all income inequality. Moreover, results appear to be unaffected by the removal of the influence of time-invariant efforts. According to his results, in Spain in 2019, the IOp accounted for approximately 73.7% of total income inequality, ranking seventh worst among European countries.

Moving beyond IOp measured with income, Rentería (2023) is, to our knowledge, the first author to provide upper bound estimates for inequality of educational opportunity. As in the aforementioned studies, he relies on the framework proposed by Niehues and Peichl (2014) and estimates two FE models: one where he fully compensates for the indirect effect of circumstances on effort and another one that offers no compensation at all. Additionally, he expands on Niehues and Peichl (2014) by assessing the role of observed time-varying circumstances on inequality of educational opportunity. Using longitudinal data from Peru, he finds that around 70% of educational

4 It is worth noting that CARRANZA (2023) employs a FE model based on data from a 3-year period. This window is notably shorter than the one used by NIEHUES and PEICHL (2014), whose average time frame covers 7 years. Nonetheless, CARRANZA (2023) offers a Monte Carlo simulation that investigates the significance of the length of the time period used to estimate a FE regression. His findings suggest that the use of a shorter time frame results in a greater number of time-invariant factors that could be classified as circumstances.

inequality, measured in standardised test scores, can be attributed to circumstances. Moreover, his findings suggest that including time-varying circumstances along with time-invariant individual fixed effects does not alter his estimates, implying that the impact of time-varying circumstances on standardised test scores is constant in practice.

3. OUR PROPOSAL

In this paper, we propose two novel aspects in the measurement of the lower and the upper bound estimates of IOp. Regarding the lower bound, we innovate by using elastic net and post-OLS elastic net estimations, which provide new estimates for IOp derived using cutting-edge machine learning techniques. Furthermore, we propose the use of additional variables not used before in the literature that potentially measure effort to discern its importance in the measurement of the upper bound of IOp.

3.1. Elastic net and post-OLS elastic net estimations for the lower bound of inequality of opportunity

We propose a novel approach to calculate lower bound estimates for IOp using elastic net and post-OLS elastic net estimations. Elastic net estimations offer a more refined solution than the lasso estimations in Hufe *et al*. (2022) by incorporating a mixture of lasso and ridge penalties. This flexibility ensures that the model benefits from both regularisation techniques, adapting to different data structures and correlation patterns resulting in improved generalisation and more accurate out-of-sample predictions, ultimately enhancing the estimation of lower bound IOp measures.

In both the elastic net and post-OLS elastic net estimation approaches, we first estimate the following model:

$$arg\,min\,\beta \sum_{i=1}^{n}\left(ln\,y_{it} - \alpha^{\text{elastic net}} - \sum_{j=1}^{k}\beta_j^{\text{elastic net}} * C_{ij}\right)^2 + \alpha\gamma\sum_{j=1}^{k}\left|\beta_j^{\text{elastic net}}\right| + .5\alpha(1-\gamma)\sum_{j=1}^{k}\left(\beta_j^{\text{elastic net}}\right)^2 \quad (1)$$

Here, α and γ are tuning parameters that control the overall penalty level and the balance between lasso and ridge penalties, respectively. We choose the optimal parameterisation of α and γ through 5-fold cross-validation.

The first lower bound estimate (LB1) uses the resulting vector $\beta_j^{\text{elastic net}}$ to construct the counterfactual distribution $\tilde{\mathrm{M}}\mathrm{t}^{\mathrm{LB1}}=(\tilde{\mu}1\mathrm{t}^{\mathrm{LB1}},\ldots,\tilde{\mu}\mathrm{it}^{\mathrm{LB1}},\ldots,\tilde{\mu}\mathrm{Nt}^{\mathrm{LB1}})$:

$$\tilde{\mu}i\,t^{LB1} = exp\left\{\hat{\alpha}^{\text{elastic net}} + \hat{\beta}^{\text{elastic net}} * Ci + \frac{\sigma^2}{2}\right\} \quad (2)$$

The second lower bound estimate (LB2) implements a post-OLS elastic net estimation. We only retain the subset $C^r \subseteq C$, i.e. those circumstances whose coefficients were not shrunk to zero in the elastic net estimation. Then, we estimate $\beta_j^{\text{elastic net}}$ by running an OLS regression on the restricted set of circumstances:

$$\ln y_{it} = \alpha^{Post\text{-}elastic\ net} + \beta^{Post\text{-}elastic\ net} * Ci^{r} + \epsilon_{it} \tag{3}$$

We use $\boldsymbol{\beta}^{Post\text{-}elastic\ net}$ to construct the counterfactual distribution $\widetilde{\mathrm{M}}\mathrm{t}^{\mathrm{LB2}} = (\tilde{\mu}1\ \mathrm{t}^{\mathrm{LB2}},\ \dots,\ \tilde{\mu}\mathrm{i}\ \mathrm{t}^{\mathrm{LB2}},\ \dots,\ \tilde{\mu}\mathrm{N}\ \mathrm{t}^{\mathrm{LB2}})$:

$$\tilde{\mu}i\ t^{LB2} = \exp\left\{\hat{\alpha}^{\text{Post-elastic net}} + \widehat{\boldsymbol{\beta}}^{\text{Post-elastic net}} * \mathbf{Ci}^{\mathbf{r}} + \frac{\sigma^2}{2}\right\} \tag{4}$$

It is important to note that LB1 and LB2 are different estimates of the same parameter vector. The choice between these two estimation methods depends on the trade-offs between prediction accuracy and model complexity. In our empirical application, we refer to the elastic net approach as our baseline LB estimate. However, we demonstrate that our main conclusions are robust to this choice.

3.2. Effort and the measurement of the upper bound of inequality of opportunity

A relevant aspect when estimating the upper bound of IOp, both for the empirical implementation but also from a philosophical point of view, is that a decision needs to be taken regarding the indirect effect of circumstances on effort that affects the outcome of interest. In the case of income, it is easy to think, for example, of a potential relationship between parental occupation and the norms regarding work effort transmitted to one's own offspring. As explained above, Niehues and Peichl (2014) explicitly model the two extreme possibilities: (1) no compensation for the indirect effects and (2) full compensation for the indirect effects. In the first approach, all the indirect effects are treated as effort and therefore are considered to be personal responsibility. Methodologically, this implies the need to account for between group differences in effort in order to obtain a direct effect of circumstances on the outcome purged from indirect effects. In the second approach, all the indirect effects are considered circumstances and therefore do not need to be separated from the direct effect of circumstances on income.

Similarly to Niehues and Peichl (2014), and in the case of no compensation, we estimate a FE model using longitudinal data by which income $(\mathrm{y}_{\mathrm{it}})$ is influenced by time-invariant observable circumstances $(\mathrm{C}_{\mathrm{i}})$ and time-varying observable effort $(\mathrm{E}_{\mathrm{it}})$. Therefore, and provided that the panel data at hand does not allow observing circumstance variables, the specification can be defined as follows:

$$\mathrm{y}_{\mathrm{it}} = \alpha_0 + \beta_0\, \mathrm{C}_{\mathrm{i}}^{\mathrm{NO}} + \eta_0\, \mathrm{E}_{\mathrm{it}} + \mu_{\mathrm{t}} + \epsilon_{\mathrm{it}} \tag{5}$$

where μ_{t} are year FE which are shifters of the income level over time which not only can account for the possibility that the effect of C_{i} on y_{it} changes over time, but also that there are circumstances beyond the individual control (such as a macroeconomic shock) that can vary. ϵ_{it} is the error term. In the final stage, and in order to obtain the upper bound of IOp with no compensation, we move from the longitudinal setting to a cross-sectional one to estimate the following reduced form:

$$\mathrm{y}_{\mathrm{it}} = \psi\, \widehat{\mathrm{C}_{\mathrm{i}}^{\mathrm{NO}}} + \upsilon_{\mathrm{it}} \tag{6}$$

which allows us to obtain the maximum extent of inequality attributable to circumstances. We then compute the parametric estimate of such distribution by replacing observed income with its prediction.

Differently, in the case of full compensation, the indirect effect of circumstances on effort that influences income is considered unfair, and arrangements need to be made to compensate individuals. This means that one needs to sterilise effort in the indirect effect. First, a FE model that does not include any effort variables allows one to obtain a measure of circumstances. That is,

$$y_{it} = u_i + u_t + \zeta_{it} \tag{7}$$

Next, we use the estimate of the estimated individual effect $\widehat{u_i}$ to sterilise all effort variables from the impact of circumstances. We do that by regressing all effort variables on the individual effect:

$$E_{it} = \kappa_1 \widehat{u_i} + u_t + e_{it} \tag{8}$$

Following from that, and in a new FE model, we use the residuals from equation (8), $\widehat{e_{it}}$ as sterilised effort variables assuming that they no longer contain the impact of circumstances. We obtain estimates of the individual effects as follows:

$$y_{it} = \lambda\widehat{e_{it}} + C_i^{FULL} + \mu_t + \xi_{it} \tag{9}$$

Again, in the final stage, and in order to obtain the upper bound of IOp with full compensation, we estimate the following reduced form:

$$y_{it} = \delta\widehat{C_i^{FULL}} + \nu_{it} \tag{10}$$

which allows us to obtain the maximum extent of inequality attributable to circumstances. We then compute the parametric estimate of such distribution by replacing observed income with its prediction.[5]

Besides the two approaches (no compensation for indirect effects and full compensation for indirect effects), we also estimate a FE model that excludes any effort variables. As Flatscher (2020) argues, if effort variables remain relatively stable over time, they can be considered circumstances rather than efforts. This benchmark estimate accounts for the direct and the indirect influence of circumstances and it is in Carranza (2023) 's notation the reduced form model. Our specification takes the following form:

$$y_{it} = \alpha_0 + \beta_0 C_i^{BENCHMARK} + \mu_t + \epsilon_{it} \tag{11}$$

where, again, (y_{it}) is our outcome of interest, income, influenced only by time-invariant observable circumstances (C_i). As in equation (6) and (10), we use the predicted FE $(\widehat{C_i^{BENCHMARK}})$ to estimate the following reduced form:

5 CARRANZA (2023) refers to the approach that completely removes the influence of effort from his measure of circumstance as the 'absolute measure of effort' and the one that considers the indirect influence of effort as a circumstance as the 'relative measure of effort'.

$$y_{it} = \psi \, C_i^{\widehat{BENCHMARK}} + \varphi_{it} \quad (12)$$

which allows us to obtain the benchmark estimates against which to compare the two other approaches.

4. DATA

To measure the effect of circumstances on income, we rely on data from the European Union – Statistics on Income and Living Conditions (EU-SILC). Our outcome of interest is household equivalised disposable income. This refers to the total income of a household that is accessible for expenditure or saving in a year, divided by the count of equivalised adults, utilising the modified OECD equivalence scale. This is the preferred outcome of most of the literature on the measurement of IOp (Ramos and Van de gaer, 2016). Our sample consists of individuals aged 25–55 with positive incomes.

The EU-SILC has two components: a cross-sectional and a longitudinal one.[6] While the cross-sectional sample collects information from respondents annually, the longitudinal sample follows each surveyed individual for four consecutive years, using a rotational scheme (Borst and Wirth, 2022).[7] Additionally, the cross-sectional sample is complemented yearly, with supplementary variables highlighting unexplored aspects of social inclusion. As Carranza (2023) points out, it is not possible to estimate both approaches using a sample simultaneously present in the cross-sectional component, and the longitudinal component because the cross-sectional sample does not allow the estimation of FE regressions and the longitudinal sample does not include retrospective information of surveyed individuals. We use the cross-sectional sample to compute the lower bound estimates of IOp and the longitudinal sample for the upper bound estimates.

Lower bound estimates are available for 2005, 2011, and 2019, as these waves include ad-hoc modules with retrospective information on the respondent's parental background and the household where she grew up. Table 1 shows the choice of circumstance variables for each lower bound estimate. The 2005 module contains relatively scarce information on the household's economic situation and the characteristics of surveyed individuals' parents. In contrast, the 2011 module includes more detailed questions on the occupation of the parents and the household composition, but, again, the information available on the household's economic circumstances is relatively scarce. Finally, in 2019, there was a substantial increase in the variables contained in the module. More questions were included about the household's economic situation, with special emphasis on material deprivation. The 2019 module also includes the size of the municipality where the surveyed individuals grew up and more detailed information on the household compo-

[6] At the time of writing, the latest release of data (provided to researchers in November 2022) includes information up to 2020 – see Table A.1 in Appendix A for all the details regarding the number of waves and the participation of each country in the cross-sectional and the longitudinal components of the EU-SILC.

[7] Complete details on the construction of the longitudinal sample can be found in Appendix B.

sition. We include all available circumstances for each specific year, grouping them as shown in Table 1.

Upper-bound estimates are available between 2005 and 2019. Given that we use a 3-year window in order to estimate the FE (we drop the reference year for which the bound is estimated), we only include individuals who have participated in all four waves of data.[8] Table A.2 in Appendix A shows the total number of observations per year used to provide upper bound estimates in Spain and the rest of European countries.

Table 1

CIRCUMSTANCE VARIABLES BY YEAR, EU-SILC, 2005, 2011 AND 2019

	2005	
Gender of respondent	**Education (both parents)** *,†	**Occupation (both parents)***
a. Male	a. No father/mother	a. No father/mother
b. Female	b. Low	b. Father/mother not working
Citizenship of respondent*	c. Medium	a. Armed forces occupations
a. Native	d. High	b. Elementary occupations
b. Migrant from EU origin	**Activity (both parents)***	c. Plant and machine operators and
c. Migrant from another region	a. No father/mother	assemblers
Loneparent family*	b. Employee	d. Craft and related trades workers
a. No	c. Self-employed	e. Skilled agricultural, forestry and fishery
b. Yes	d. Unemployed	workers
Number of siblings*	e. Retired	f. Service and sales workers
a. None	f. Housework	g. Clerical support workers
b. 1	g. Other inactive	h. Technicians and associate professionals
c. 2		i. Professionals
d. 3		j. Managers
e. 4		
f. 5 or more		

8 As noted by NIEHUES and PEICHL (2014) and CARRANZA (2023), we employ previous years' data to estimate the fixed effects. For example, when estimating IOp in 2019, we utilize data from 2018, 2017, and 2016 to derive the fixed effects. It should be noted that it is not feasible to employ the preceding years for estimating IOp in 2005, 2006, and 2007, since the EU-SILC contains data from 2004 onwards. Instead, and following CARRANZA (2023), we use the subsequent three years. Thus, for estimating IOp in 2008, we rely on data from 2009, 2010, and 2011 to calculate the fixed effects.

	2011	
Gender of respondent	**Financial situation of the**	**Activity (both parents)** *
a. Male	**household** *,†	a. No father/mother
b. Female	a. Very bad	b. Employee
Migrant	b. Bad	c. Self-employed
a. Native	c. Moderately bad	d. Unemployed
b. Migrant from EU origin	d. Moderately good	e. Retired
c. Migrant from another region	e. Good	f. Housework
Monoparental family*	f. Very good	g. Other inactive
a. No	**Country of birth (both**	**Number of working people***
b. Yes	**parents)** *	a. None
Number of siblings*	a. Unknown	b. 1
a. None	b. Present country	c. 2
b. 1	c. Other EU-27	d. 3 or more
c. 2	d. Other Europe	**Occupation (both parents)** *
d. 3	e. Outside Europe	a. No father/mother
e. 4	**Citizenship (both parents)** *	b. Father/mother not working
f. 5 or more	a. Unknown	a. Armed forces occupations
Tenancy status*	b. Present country	b. Elementary occupations
a. Tenant	c. Other EU-27	c. Plant and machine operators and
b. Free accommodation	d. Other Europe	assemblers
c. Owner	e. Outside Europe	d. Craft and related trades workers
	Education (both parents) *,†	e. Skilled agricultural, forestry and fishery workers
	a. No father/mother	f. Service and sales workers
	b. Low	g. Clerical support workers
	c. Medium	h. Technicians and associate professionals
	d. High	i. Professionals
		j. Managers
		Managerial status (both parents) *
		a. Supervisor
		b. Non-supervisor

2019		
Gender of respondent	**Basic school needs** *,†	**Activity (both parents)** *
a. Male	a. No	a. No father/mother
b. Female	b. Yes	b. Employee
Migrant	**Daily protein** *,†	c. Self-employed
a. Native	a. No	d. Unemployed
	b. Yes	e. Retired
b. Migrant from EU origin	**At least one week for**	f. Housework
c. Migrant from another region	**holidays per year** *,†	g. Other inactive
Loneparent family*	a. No	**Number of working people***
a. No	b. Yes	a. None
b. Yes	**Country of birth (both**	b. 1
Number of siblings (adults) *	**parents)** *	c. 2
a. None	a. Unknown	d. 3 or more
b. 1	b. Present country	**Occupation (both parents)** *
c. 2	c. Other EU-27	a. No father/mother
d. 3	d. Other Europe	b. Father/mother not working
e. 4	e. Outside Europe	a. Armed forces occupations
f. 5 or more	**Citizenship (both parents)** *	b. Elementary occupations
Tenancy status*	a. Unknown	c. Plant and machine operators and
a. Tenant	b. Present country	assemblers
b. Free accommodation	c. Other EU-27	d. Craft and related trades workers
c. Owner	d. Other Europe	e. Skilled agricultural, forestry and fishery workers
Degree of urbanisation*	e. Outside Europe	f. Service and sales workers
a. Rural area	**Education (both parents)** *,†	g. Clerical support workers
b. Town or suburb	a. No father/mother	h. Technicians and associate professionals
c. City	b. Low	i. Professionals
Financial situation of the	c. Medium	j. Managers
household *,†	d. High	**Managerial status (both parents)** *
a. Very bad		a. Supervisor
b. Bad		b. Non-supervisor
c. Moderately bad		
d. Moderately good		
e. Good		
f. Very good		

Note: All variables marked with * are from the retrospective ad-hoc modules. For variables marked with †, we compute a new variable that registers the average cohort outcome in the region where the respondent was born, e.g., the average mother's education in a given region for a given cohort.

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Finally, in Tables A.3 to A.5 in Appendix A we show summary statistics for the 2005, 2011 and 2019 cross-sectional sample for the set of circumstances that capture retrospective information on surveyed individuals' parents and household characteristics, as well as some individual characteristics used in our estimations. For example, panel A of Table A.5 shows that, on average, cross-sectional surveyed individuals are 41 years old in 2019. About 51% of the sample are females. Also 13% of the adults considered come from an immigrant origin. Regarding the level of education, a large number of individuals have acquired at most (upper) secondary and post-secondary non-tertiary education. Also, the majority of individuals are employed.

Regarding the cross-sectional retrospective information on surveyed individuals, Panel B shows that 11% of individuals grew up in a single-parent household. Individuals' fathers, on average, showed a higher level of education than individual's mothers and most of the surveyed individuals' parents are not of immigrant origin. Again, individual's fathers have, on average, higher employment rates than mothers and hold more often managerial positions. As to the household economic situation, most individuals grew up in a household whose financial situation was moderately good. Yet, approximately, 5% of the sample could not afford basic school needs, and 8% could not afford a meal with meat, chicken, fish (or their vegetarian equivalent) daily. Finally, 28% of the sample could not afford one week of annual holiday away from home.

5. RESULTS

In this section, we provide our main results. First, we present our estimates for the lower bound of IOp. Next, we continue with the results for the upper bound. Each subsection starts with the results relative to Spain and then compares them with the rest of the European countries.

5.1. Lower bound

Table 2 shows the lower bound estimates of IOp in 2005, 2011 and 2019 for Spain. We find that the lower bound of IOp has steadily increased from explaining 11.5% of total inequality in 2005 to accounting for 17.57% in 2019, which represents a substantive increase over the period. In terms of Gini, it is equivalent to an evolution from 37% to 45%.[9] Overall, this suggests that unfair inequalities have become more relevant for explaining income inequality in Spain during this period.

9 We base our main analysis on the MLD in order to ease the comparison vis-à-vis the academic literature on IOp.

Table 2

LOWER BOUND ESTIMATES OF IOP USING ELASTIC NET, SPAIN, 2005, 2011 AND 2019

Year	MLD	IOp	IOR
2005	0.194	0.022	11.15%
2011	0.197	0.026	13.35%
2019	0.205	0.036	17.57%

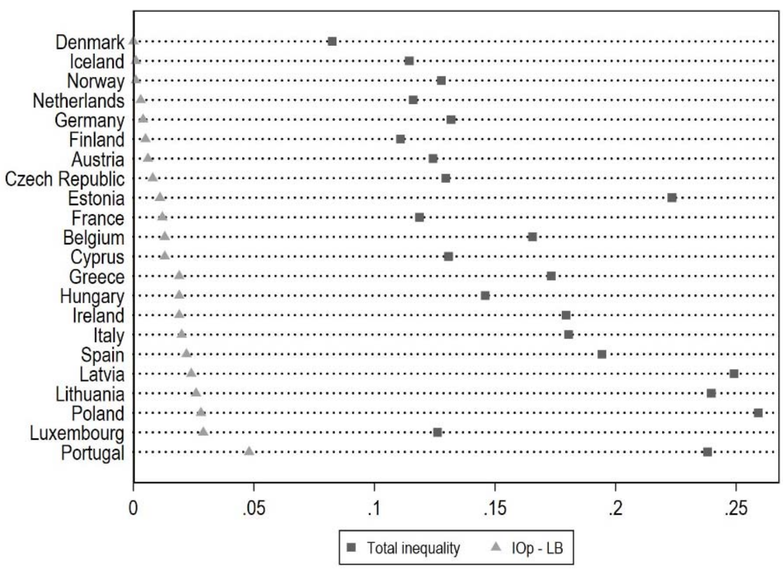
Note: IOp refers to inequality of opportunity, and IOR indicates the ratio of inequality of opportunity to total inequality.

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Figures 1, 2, and 3 show the lower bound estimates across European countries for 2005, 2011 and 2019, respectively. We plot in each figure the estimated levels of lower bound inequality of opportunity and total income inequality per country. In 2005 and 2011, IOp estimates across European countries remain relatively stable, ranging from about 0.001 to 0.05. In 2019, however, we document an upsurge in inequality of opportunity across European countries: the lower bound estimates for that year range from 0.002 to 0.077. We saw in Table 2 that Spain has experienced a persistent rise in inequality of opportunity. It is also clear from Figures 1, 2, and 3 that Spain has also worsened its relative position vis-à-vis the rest of European countries, moving from the sixth to the fourth country in Europe with the highest level of IOp.

Figure 1

LOWER BOUND ESTIMATES, EUROPE, 2005

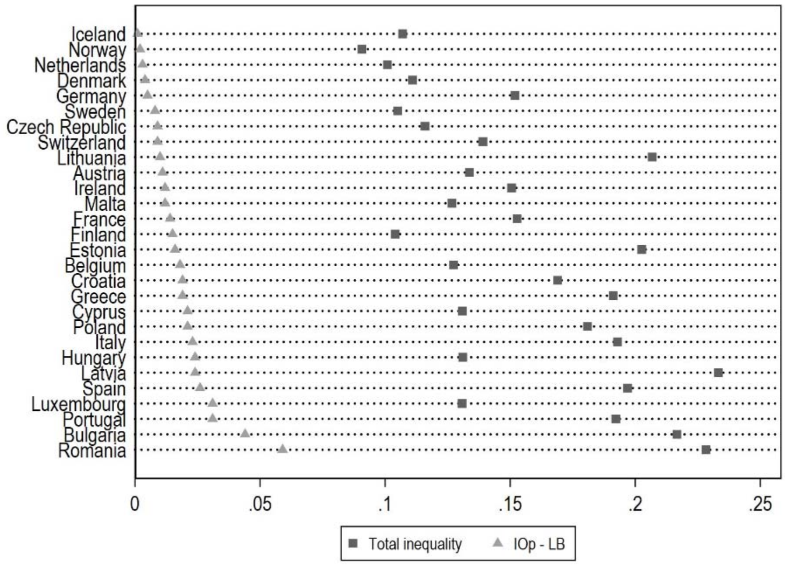


Note: Data for Sweden, Slovenia, Slovakia, and the UK is not available.

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Figure 2

LOWER BOUND ESTIMATES, EUROPE, 2011

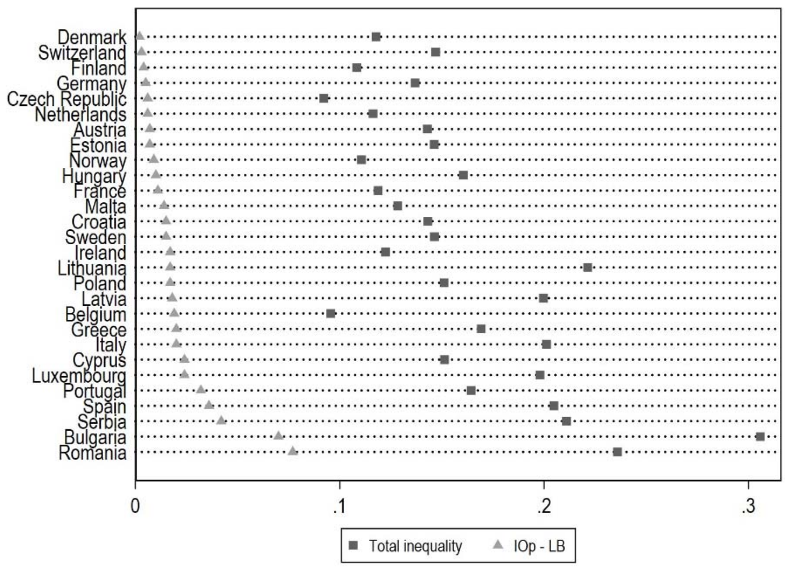


Note: Data for Slovenia, Slovakia, and the UK is not available.

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Figure 3

LOWER BOUND ESTIMATES, EUROPE, 2019



Note: Data for Slovenia, and Slovakia is not available.

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Brunori *et al.* (2023) show through rank correlations that existing literature on inequality of opportunity in Europe is not internally consistent and only moderately consistent with IGE rankings, being their forest approach a notable exception. We update this exercise to test how reliable the novel elastic net method is. Table 3 shows that our baseline elastic net estimates strongly correlate with Brunori *et al*. (2023) 's forest estimates (rank correlation at 0.94) based on the Gini index. The baseline elastic net estimates are also aligned with the CV approach in Brunori *et al.* (2019) – with a rank correlation of 0.89. Carranza (2023) 's MLD-based estimates appear to be more inconsistent with both elastic net and CV. Our elastic net estimates are closely aligned with IGE, with a rank correlation of 0.80. The post-OLS elastic net estimates are highly correlated with our baseline elastic net model yet relatively less consistent with the remaining approaches from the literature and IGE estimates. This mirrors the findings of Hufe *et al*. (2022) for lasso: their post-OLS lasso estimates are also less internally consistent than their standard lasso IOp.

Table 3

RANK CORRELATIONS OF EXISTING STUDIES

	Existing studies			**This paper**	
	Brunori *et al.* (2019)	**Carranza *et al*. (2023)**	**Brunori *et al.* (2023)**	**Elastic Net**	**Elastic Net (post OLS)**
Panel (a): Equality of opportunity in 2011 (MLD)					
Elastic Net				1.000	
Elastic Net (post OLS)				0.875	1.000
Brunori *et al.* (2019)	1.000			0.893	0.773
Carranza *et al.* (2023)	0.629	1.000		0.570	0.385
Panel (b): Equality of opportunity in 2011 (Gini)					
Brunori *et al.* (2023)			1.000	0.939	0.773
Panel (c): Equality of opportunity in 2019 (MLD)					
Carranza *et al.* (2023)		1.000		0.735	0.765
Panel (d): Intergenerational Elasticity (9 countries)					
Stuhler (2018) & Carmichael *et al.* (2020)			0.833	0.800	0.617

Note: This table shows country rank correlations in inequality of opportunity estimates across existing studies. Panel (a) is based on the intersection of countries included in this paper across studies that provide MLD-based IOp estimates for 2011. Panel (b) is based on the intersection of countries included in this paper across studies that provide Gini-based IOp estimates for 2011. Panel (c) is based on the intersection of countries included in this paper across studies that provide MLD-based IOp estimates for 2019. Ranks in STUHLER (2018) and CARMICHAEL *et al*. (2020) are calculated from consensus estimates of the intergenerational earnings elasticity (IGE). All rank correlations are based on Spearman's *r*.

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC), and the estimates from the included studies.[10]

[10] We thank Paolo BRUNORI, Paul HUFE, and Daniel MAHLER for sharing the IGE data with us.

5.2. Upper bound

In this section, we present the results of the upper bound estimation as explained in section 3.2 while following the seminal work by Niehues and Peichl (2014) and, in comparison with the most recent estimates for Europe by Carranza (2023). The second column of Table 4 presents our inequality estimates for Spain between 2005 and 2019 using the MLD. Column 3 shows the results of the reduced form model (equations 11–12 in this paper) – used as a benchmark here – which accounts for the direct and indirect influence of circumstances. Columns 4 and 5 explicitly account for effort. In column 4, we attempt to completely remove the influence of effort from circumstances (equations 5–6) which is referred as the 'no compensation approach' in Niehues and Peichl (2014). We do that by considering the following effort variables: number of hours worked per week, type of contract, years in paid work relative to age and the year of first job relative to age. Column 5 instead provides the results when the indirect effect of effort is considered a circumstance – termed by the same authors as the 'full compensation' approach. Columns 6 to 8 provide the same results but as a ratio of total inequality.

Table 4

UPPER BOUND ESTIMATES FOR IOP, SPAIN, 2005 – 2019

Year (1)	MLD (2)	IOp benchmark (3)	IOp no compensation (4)	IOp full compensation (5)	IOR benchmark (6)	IOR no compensation (7)	IOR full compensation (8)
2005	0.20	0.09	0.08	0.09	43.40	41.88	42.01
2006	0.19	0.09	0.08	0.08	45.69	43.84	43.97
2007	0.19	0.08	0.07	0.07	41.91	39.74	39.81
2008	0.17	0.08	0.08	0.08	46.91	45.86	45.95
2009	0.20	0.09	0.08	0.08	43.16	38.92	39.12
2010	0.21	0.09	0.08	0.09	45.67	41.13	41.19
2011	0.20	0.09	0.08	0.08	45.17	41.76	41.82
2012	0.21	0.10	0.09	0.09	45.51	42.69	42.77
2013	0.20	0.10	0.09	0.09	49.94	44.22	44.54
2014	0.23	0.10	0.09	0.10	43.89	40.92	41.07
2015	0.22	0.13	0.12	0.12	59.90	55.45	55.76
2016	0.20	0.14	0.13	0.13	70.91	65.67	65.80
2017	0.22	0.14	0.14	0.14	63.55	63.44	63.44
2018	0.17	0.13	0.12	0.12	75.76	70.91	70.92
2019	0.20	0.14	0.13	0.13	67.98	67.13	67.15

Note: IOp refers to inequality of opportunity, and IOR indicates the ratio of inequality of opportunity to total inequality.

Source: Authors' computation, using data from the longitudinal component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Two important findings emerge from Table 4. First, results show a substantial increase in the upper bound of IOp in Spain from 2005 to 2019 regardless of the model used. While in 2005, the estimates of the reduced form model (Column 6) indicate that about 43.4% of the inequality in Spain was explained by IOp, the same figure in 2019 stands at 68%. This is an increase of 24.6 percentage points. The rise in IOp seems unrelated to the business cycle as we observe higher figures in recent years when the Spanish economy has been growing than in the aftermath of the 2008 financial crisis. That is, IOp is becoming a structural problem in Spain that does not adjust even when individuals may find more opportunities in a growing economy.

Second, and particularly relevant from a methodological point of view, results indicate that the overall impact of the effort variables we can use in our models is very limited. The upper bound estimates of IOp do not change when we remove the influence of effort from circumstances. Note the similarity between Columns 4 and 5 and between 7 and 8. The largest difference in the results (even if small) is between the benchmark estimates and the other two estimates, which confirms previous findings in the literature that the indirect influence of circumstances on efforts does not make any difference for the upper bound estimate of IOp (Niehues and Peichl, 2014; Carranza, 2023). This is true at least for the type of effort variables contained in the surveys available to researchers nowadays.

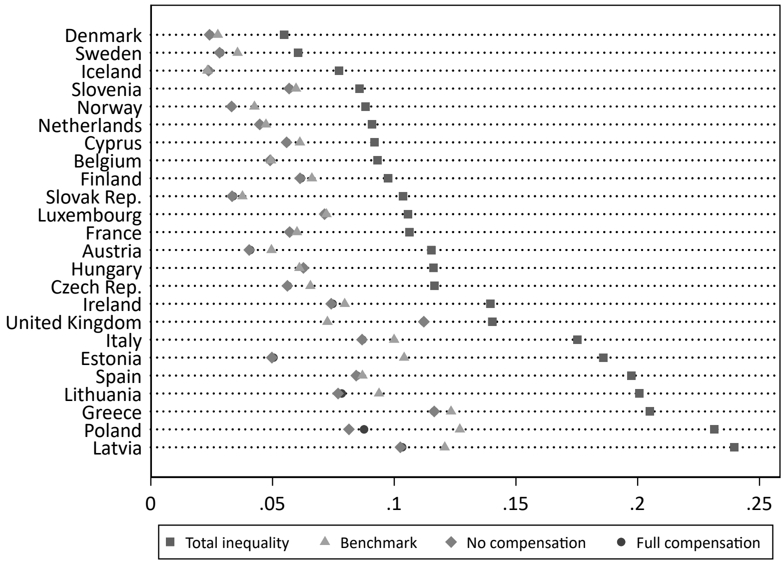
Figures 4, 5 and 6 present upper bound estimates across European countries for 2005, 2011 and 2019, respectively. In each of the three figures, we plot total inequality as well as the three estimates: (1) 'no compensation' for the indirect effects of effort on circumstances, (2) 'full compensation' for the indirect effects of effort on circumstances, and (3) exclusion of any effort variables – referred as 'benchmark'.

Between 2005 and 2019, total inequality has increased in Cyprus, Denmark, Hungary, Italy, Lithuania, Luxembourg, the Netherlands, Sweden and Slovenia. However, it is noteworthy that the increase in IOp has been even more pronounced since most of the European countries considered, except Greece, Hungary, Luxembourg and Sweden, have experienced a rise in the IOR during this period.

Consistent with the findings presented in Table 4 for Spain, compensating for the indirect effect of circumstances through effort on income and no compensating at all yield similar results in all the countries considered. IOp remains unchanged even when the indirect influence of circumstances on effort is removed from the analysis. Nevertheless, despite being small in magnitude, noticeable differences can be observed between the benchmark estimate and both the estimates with no compensation and full compensation, with the benchmark estimate being higher.

Figure 4

UPPER BOUND ESTIMATES, EUROPE, 2005

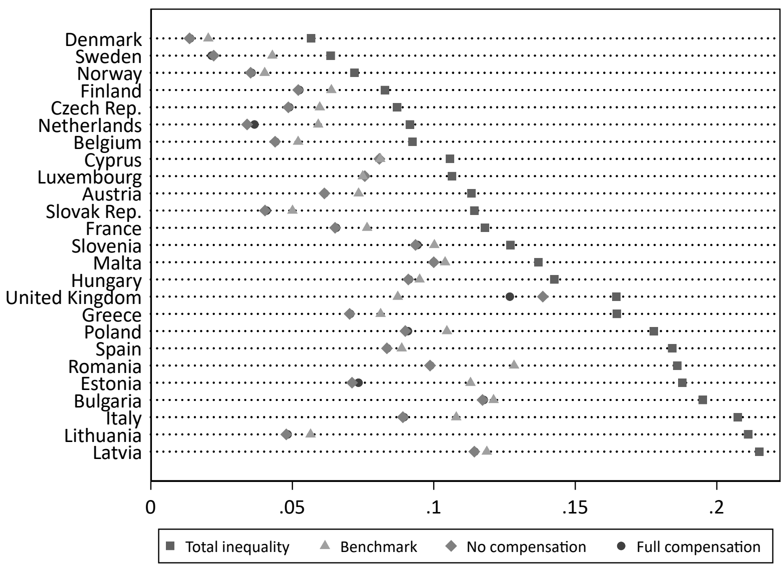


Note: Data for Bulgaria, Switzerland, Germany, Croatia, Malta, Romania and Serbia is not available. Results for Portugal are omitted due to inconsistencies.

Source: Authors' computation, using data from the longitudinal component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Figure 5

UPPER BOUND ESTIMATES, EUROPE, 2011

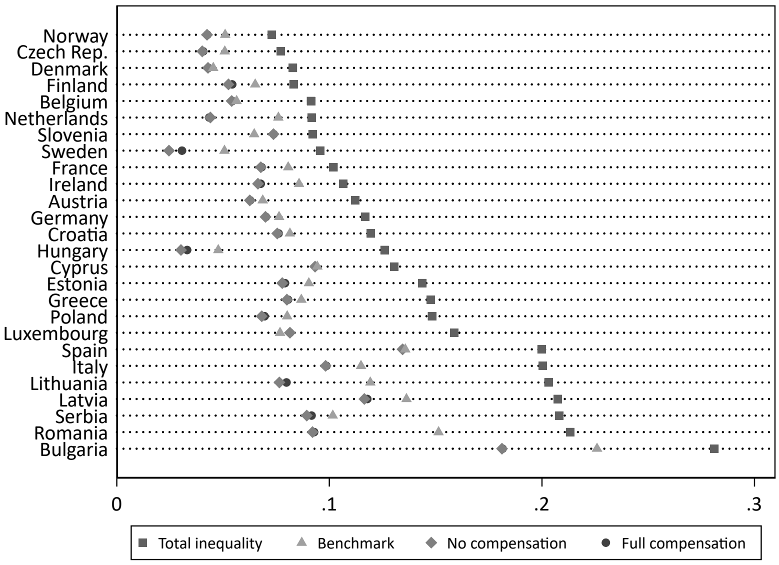


Note: Data for Switzerland, Croatia, Ireland, Iceland and Serbia is not available. Results for Portugal are omitted due to inconsistencies.

Source: Authors' computation, using data from the longitudinal component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Figure 6

UPPER BOUND ESTIMATES, EUROPE, 2019



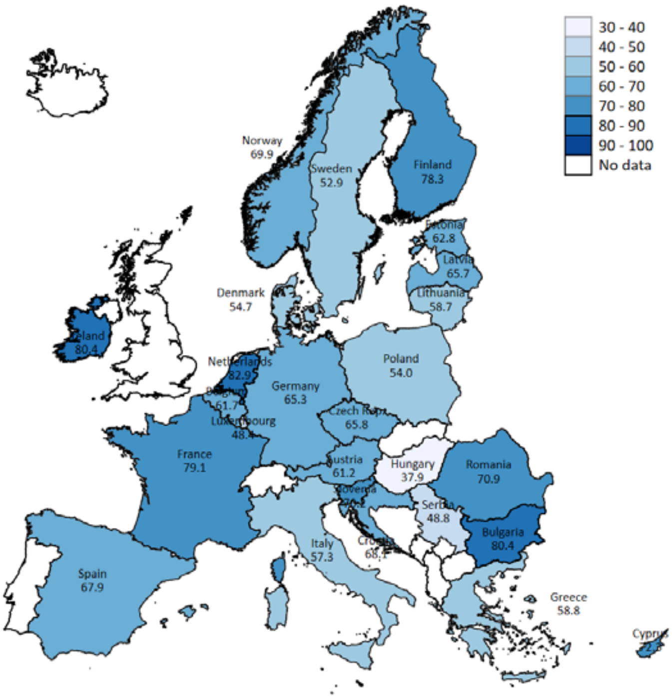
Note: Data for Switzerland, Iceland, Malta, Slovakia, and the United Kingdom is not available. Results for Portugal are omitted due to inconsistencies.

Source: Authors' computation, using data from the longitudinal component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Figure 7 displays the proportion of total inequality attributed to IOp across European countries in 2019. In this case, the benchmark estimates are employed as the measure of IOp. Remarkably, Spain occupies the eleventh position in terms of the proportion of IOp on total inequality, aligning with Carranza's (2023) results. However, when effort is taken into consideration (either no or full compensation), the ranking shifts, with Spain emerging as the third country with the highest IOp, alongside Cyprus and Slovenia

Figure 7

INEQUALITY OF OPPORTUNITY OVER TOTAL INEQUALITY, EUROPE, 2019



Note: IOp benchmark estimates are displayed. Data for Switzerland, Iceland, Malta, Slovakia, and the United Kingdom is not available. Results for Portugal are omitted due to data inconsistency.

Source: Authors' computation, using data from the longitudinal component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

6. CONCLUSIONS

This paper provides new estimates for IOp in Spain between 2005 and 2019 using data from the EU-SILC. On the one hand, we propose the estimation of the lower bound of IOp using a new machine learning technique – elastic net and post-OLS elastic net– that provides a higher level of flexibility in comparison with previously used similar techniques and that allows for improved generalisation and more accurate out-of-sample predictions. On the other hand, we provide a careful estimation of the upper bound of IOp while considering different approaches regarding the role of effort. Results allow for the analysis of trends over a period of 14 years and also the comparison of the Spanish case against the rest of European countries.

Lower bound estimates indicate that Spain exhibits an IOR of 17%. Spain has also experienced a remarkable increase in IOR during the period of analysis, moving from 11% in 2005 to 17% in 2019. As a result, Spain deteriorated its relative position in terms of unfair inequality, moving from the sixth position in 2005 to being in 2019 the fourth country in Europe with the highest lower bound estimate for IOp. From a methodological standpoint, our novel use of the elastic net algorithm aligns strongly with both established approaches for measuring IOp and IGE rankings.

Upper bound estimates indicate that Spain exhibits an IOR of 68%. Between 2005 and 2019, there has been a significant increase of 24.6 percentage points in IOR, showing a greater rise in IOp than overall inequality. In the European context, Spain is ranked eleventh in terms of benchmark estimates for IOR. However, when factoring in the effect of effort (either no or full compensation), Spain emerges as the third-ranked country, following Cyprus and Slovenia. From a methodological point of view, we confirm previous results in the literature in the sense that accounting for or ignoring the indirect effects of circumstances on effort yields similar results.

Our findings suggest several policy recommendations for addressing intergenerational poverty and inequality of opportunity in Spain. Firstly, we recommend investing in public policies that improve socioeconomic conditions during childhood. It would be beneficial to initiate a tax reform in line with the recommendations of the White Paper on Tax Reform (Comité de Personas Expertas para elaborar el Libro Blanco sobre la Reforma Tributaria, 2022). This reform would require an increase in indirect revenue by expanding the bases of the welfare state for more ambitious policies. Spain's implicit VAT rates place it at the end of the ranking of European Union countries, only surpassed by Romania and Italy (Comité de Personas Expertas para elaborar el Libro Blanco sobre la Reforma Tributaria, 2022: 151). A gradual transition to a simplified VAT applying a single rate to a broader tax base is recommended. As this is a regressive policy, it is important to offset households with lower income, ensuring that the tax system and direct benefits maintain their progressivity (Comité de Personas Expertas para elaborar el Libro Blanco sobre la Reforma Tributaria, 2022: 158-162). This reform would bring us closer to the equity-generating capacity of the social benefit systems of the EU-8 countries. Secondly, public policies need to be more incisive in the area of family policies. We recommend the creation of a universal, refundable tax deduction of 100 euros per month per dependent child under 18 years of age in Spain, aligning us closer to the average of these policies in the European Union (Ibarra *et al*., 2021). This policy would increase the system's progressivity, supporting those who do not benefit from the current minimum for descendants (Zalakain, 2019). A recent literature review shows that this measure would positively impact child poverty and improve child well-being from an educational and health perspective (Zalakain, 2021).

Thirdly, implementing the Minimum Vital Income (IMV) needs to be improved. The IMV has managed to improve the level of coverage that regional incomes had before its introduction (Dirección General de Diversidad Familiar, 2019). However, three years after its launch, the coverage objective of 850,000 households has barely been reached. It is necessary to streamline the application process, reduce the administrative burden, and provide resources for denied applications (EAPN, 2021). Lastly, we believe that inequality of wealth should be reduced. Recent literature suggests that wealth inequality has a marked component of inequality of opportunities largely explained by the role of inheritances and donations (Palomino *et al*., 2021; Salas-Rojo and Rodríguez, 2022). The White Paper discusses the issue of harmonising wealth taxation, highlighting the downward competition among Autonomous Communities in the context of wealth and inheritance taxes (Blanchard *et al*., 2021: 72-73). To emphasise the redistributive nature of the tax and boost equality of opportunities, the Blanchard-Tirole commission suggests directly transferring the revenues from these taxes to, firstly, a system of individual accounts for the most vulnerable people and, secondly, early-stage childhood programmes. Implementing these recommendations would put Spain on a path towards reducing intergenerational poverty and enhancing equality of opportunities.

References

BALCÁZAR, C. F. (2015): “Lower bounds of inequality of opportunity and measurement error”, *Economics Letters*, 137: 102–105.

BLANCHARD, O.; J.; TIROLE, C.; GOLLIER, M.; REGUANT, D.; RODRIK, S.; STANTCHEVA, A.; BÖRSCH-SUPAN, C.; DIEHL and C. PROPPER (2021): “Major future economic challenges”, France Stratégie.

BORST, M. and H. WIRTH (2022): “EU-SILC Tools: eusilcpanel_2020 - first computational steps towards a cumulative sample based on the EU-SILC longitudinal datasets; Update”, *GESIS Papers*, 2022/10.

BOURGUIGNON, F.; F. H. G. FERREIRA and M. MENÉNDEZ (2007): “Inequality of Opportunity in Brazil”, *Review of Income Wealth*, 53(4): 585–618.

BRUNORI, P.; P. HUFE and D. MAHLER (2023): “The roots of inequality: Estimating inequality of opportunity from regression trees and forests”, *Scandinavian Journal of Economics*, forthcoming.

BRUNORI, P.; V. PERAGINE and L. SERLENGA (2019): “Upward and downward bias when measuring inequality of opportunity”, *Social Choice and Welfare*, 52: 635–661.

BRZEZINSKI, M. (2020): “The evolution of inequality of opportunity in Europe”, *Applied Economics Letters*, 27(4): 262–266.

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.

CARRANZA, R. (2023): “Upper and lower bound estimates of inequality of opportunity: A cross-national comparison for Europe”, *Review of Income and Wealth,* forthcoming.

COMITÉ DE PERSONAS EXPERTAS PARA ELABORAR EL LIBRO BLANCO SOBRE LA REFORMA TRIBUTARIA (2022): *Libro Blanco sobre la Reforma Tributaria,* Instituto de Estudios Fiscales.

DIRECCIÓN GENERAL DE DIVERSIDAD FAMILIAR Y SERVICIOS SOCIALES (2019): *Informe de rentas mínimas de inserción. Año 2019,* Ministerio de Derechos Sociales y Agenda 2030.

EAPN (2021): “El ingreso mínimo vital un año después. La perspectiva autonómica”, *Publicaciones EAPN.*

EUROSTAT (2022): “Microdatasets availability table”. Retrieved October 26, 2022, from https://ec.europa.eu/eurostat/documents/203647/771732/Datasets-availability-table.pdf

FERREIRA, F. H. G. and J. GIGNOUX (2011): “The measurement of inequality of opportunity: theory and an application to Latin America”, *Review of Income and Wealth*, 57(4): 622–657.

FILAURO, S.; F. PALMISANO and V. PERAGINE (2023): “The Evolution of Inequality of Opportunity in Europe”. ECINEQ *Working Paper Series,* 2023-644.

FLATSCHER, D. (2020): “How uneven is the playing field? Inequality of socioeconomic opportunity in the UK, 2009-2017”. LSE Department of Social Policy, *Working Paper Series*, 06-20.

FLEURBAEY, M. (1994): “On fair compensation”, *Theory and decision*, *36*(3): 277–307.

Hothorn, T.; K. Hornik and A. Zeileis (2006): "Unbiased recursive partitioning: A conditional inference framework", *Journal of Computational and Graphical Statistics,* 15(3): 651–674.

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: 395–427.

Hufe, P.; A. Peichl; J. E. Roemer and M. Ungerer (2017): "Inequality of Income acquisition: The role of childhood circumstances", *Social Choice and Welfare* 143 (3–4): 499–544.

Iacovou, M. and P. Lynn (2013): "Implications of the EU-SILC following rules, and their implementation, for longitudinal analysis", Institute for Social and Economic Research (ISER), *Working Paper Series*, 2013-17.

Ibarra, R.; A. E. Terán and G. González-Bueno (2021): *Por una prestación para la crianza. Escenarios para una deducción fiscal reembolsable en España.* Save the Children.

Marrero, G. A. and J. G. Rodríguez (2012): "Inequality of opportunity in Europe", *Review of Income and Wealth*, *58*(4): 597–621.

Niehues, J. and A. Peichl (2014): "Upper bounds of inequality of opportunity: theory and evidence for Germany and the US", *Social Choice and Welfare*, 43: 73–99.

Palomino, J. C.; G. A. Marrero; B. Nolan and J. G. Rodríguez (2021): "Wealth inequality, intergenerational transfers, and family background", *Oxford Economic Papers*, 74(3): 643–670.

Palomino, J. C.; G. A. Marrero and J. G. Rodríguez (2019): "Channels of inequality of opportunity: The role of education and occupation in Europe", *Social Indicators Research*, 143(3): 1045–1074.

Ramos, X. and D. Van de gaer (2016): "Approaches to inequality of opportunity: principles, measures and evidence", *Journal of Economic Surveys*, 30: 855-883.

Rentería, J. M. (2023): "Inequality of educational opportunity and time-varying circumstances: Longitudinal evidence from Peru", *The Journal of Development Studies*, 59(2): 258-278.

Roemer, J. E. (1993): "A pragmatic theory of responsibility for the egalitarian planner", *Philosophy & Public Affairs*, 146–166.

Roemer, J. E. (1998): *Equality of opportunity,* Harvard University Press.

Salas-Rojo, P. and J. G. Rodríguez (2022): "Inheritances and wealth inequality: a machine learning approach", *The Journal of Economic Inequality*, 20: 27–51.

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

Zalakain, J. (2019): "La fiscalización de las políticas sociales: funcionamiento e impacto de las deducciones fiscales reembolsables", *Zerbitzuan: Gizarte Zerbitzuetarako Aldizkaria = Revista de Servicios Sociales*, 68: 111–141.

Zalakain, J. (2021): "Diseño e impacto de las prestaciones por hijos a cargo en los países de la OCDE", *Zerbitzuan: Gizarte Zerbitzuetarako Aldizkaria = Revista de Servicios Sociales*, (74): 5-32.

APPENDIX A

Table A.1

DATA AVAILABILITY BY COUNTRY AND YEAR, CROSS-SECTIONAL AND LONGITUDINAL FILES

		AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	CH	IS	NO	RS	UK
2020	Long	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	
2019	Long	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X	X	
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	
2018	Long	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2017	Long	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2016	Long	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2015	Long	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2014	Long	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2013	Long	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X		X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2012	Long	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
2011	Long	X	X	X	X	X		X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X		X	X		X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
2010	Long	X	X	X	X	X		X	X	X	X	X	X		X		X	X	X	X	X	X	X	X	X	X	X	X		X	X		X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
2009	Long	X	X	X	X	X		X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X		X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
2008	Long	X	X	X	X	X		X	X	X	X	X	X		X	X	X	X	X	X		X	X	X	X	X	X	X		X	X		X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
2007	Long	X	X	X	X	X		X	X	X	X	X	X		X	X	X	X	X	X		X	X	X		X	X	X		X	X		X
	Cross	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X
2006	Long	X	X		X	X		X	X	X	X	X	X		X	X	X	X	X	X		X	X	X		X	X	X		X	X		X
	Cross	X	X		X	X	X	X	X	X	X	X	X		X	X	X	X	X	X		X	X	X		X	X	X		X	X		X
2005	Long	X	X						X	X	X	X	X			X	X		X					X		X	X	X		X	X		X
	Cross	X	X		X	X	X	X	X	X	X	X	X		X	X	X	X	X	X		X	X	X		X	X	X		X	X		X
2004	Cross	X	X					X	X	X	X	X	X			X	X		X					X		X				X	X		

Source: Eurostat (2022).

Table A.2

TOTAL NUMBER OF OBSERVATIONS PER YEAR, EUROPE, 2005, 2011 AND 2019

Year	Number of observations		
	2005	2011	2019
AT	1055	1013	1019
BE	1081	827	1522
BG		1300	2424
CY	862	639	870
CZ	2943	1335	1477
DE			1425
DK	271	1163	489
EE	429	1033	1093
EL	1086	1309	3339
ES	2742	3263	1894
FI	1355	1081	1550
FR	4289	4679	3466
HR			1373
HU	1518	1557	924
IE	381		420
IS	569		
IT	4063	3528	3522
LT	952	1332	815
LU	2559	2687	816
LV	677	1084	885
MT		566	
NL	2437	1803	1722
PL	3415	2898	2112
PT		1296	3585
RO		1736	1491
SE	1241	941	654
SI	1920	2253	1629
SK	1524	2073	
CH			
IS			
NO	2511	1876	739
RS			1227
UK	1344	1023	

Source: Authors' computation, using data from the longitudinal component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Table A.3

SUMMARY STATISTICS, 2005

	Mean/Prop.	Std. Dev.
Panel A: Individual and household characteristics		
Age	40,698	8,743
Female	0.515	0.500
Migrant status		
Native	0.918	0.274
Migrant from EU origin	0.025	0.157
Migrant from another region	0.057	0.231
Level of education		
Primary	0.234	0.423
Secondary	0.509	0.500
Tertiary	0.257	0.437
Labour market status		
Employed	0.761	0.427
Unemployed	0.077	0.267
Student	0.021	0.145
Other inactive	0.140	0.347
Panel B: Retrospective information on surveyed individuals		
Loneparent family	0.090	0.287
Number of siblings		
0	0.113	0.316
1	0.324	0.468
2	0.245	0.430
3	0.134	0.341
4	0.074	0.262
5 or more	0.110	0.312
Highest level of education attained (father)		
No father	0.072	0.258
Low	0.411	0.492
Medium	0.417	0.493
High	0.100	0.300
Highest level of education attained (mother)		
No mother	0.026	0.158
Low	0.484	0.500
Medium	0.425	0.494
High	0.065	0.246
Labour market status (father)		
No father	0.064	0.244
Other inactive	0.018	0.133
Housework	0.003	0.056
Retired	0.016	0.127
Unemployed	0.006	0.078
Self-employed	0.204	0.402
Employed	0.689	0.463

Labour market status (mother)		
No mother	0.021	0.142
Other inactive	0.014	0.116
Housework	0.429	0.495
Retired	0.011	0.106
Unemployed	0.003	0.055
Self-employed	0.086	0.281
Employed	0.436	0.496
Main occupation (father)		
No father	0.070	0.2544
Father not working	0.023	0.150
Armed forces occupation	0.013	0.113
Elementary occupations	0.107	0.309
Plant and machine operators and assemblers	0.132	0.338
Craft and related trades workers	0.232	0.422
Skilled agricultural, forestry and fish	0.127	0.332
Service and sales workers	0.046	0.209
Clerical support workers	0.047	0.212
Technicians and associate professionals	0.071	0.258
Professionals	0.061	0.239
Managers	0.073	0.260
Main occupation (mother)		
No mother	0.022	0.145
Mother not working	0.354	0.478
Armed forces occupation	0.000	0.016
Elementary occupations	0.135	0.342
Plant and machine operators and assemblers	0.043	0.203
Craft and related trades workers	0.065	0.247
Skilled agricultural, forestry and fish	0.092	0.289
Service and sales workers	0.086	0.281
Clerical support workers	0.064	0.244
Technicians and associate professionals	0.063	0.243
Professionals	0.047	0.212
Managers	0.028	0.166

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Table A.4

SUMMARY STATISTICS, 2011

	Mean/Prop.	Std. Dev.
Panel A: Individual and household characteristics		
Age	41,186	8,789
Female	0.515	0.500
Migrant status		
Native	0.899	0.301
Migrant from EU origin	0.038	0.191
Migrant from another region	0.063	0.243
Level of education		
Primary	0.197	0.398
Secondary	0.503	0.500
Tertiary	0.300	0.458
Labour market status		
Employed	0.767	0.422
Unemployed	0.090	0.287
Student	0.021	0.144
Other inactive	0.121	0.326
Loneparent family	0.098	0.297
Panel B: Retrospective information on surveyed individuals		
Loneparent family	0.098	0.297
Number of siblings		
0	0.014	0.118
1	0.255	0.436
2	0.396	0.489
3	0.194	0.395
4	0.077	0.266
5 or more	0.065	0.245
Highest level of education attained (father)		
No father	0.007	0.080
Low	0.573	0.495
Medium	0.297	0.457
High	0.123	0.329
Highest level of education attained (mother)		
No mother	0.002	0.043
Low	0.631	0.482
Medium	0.282	0.450
High	0.085	0.279

Labour market status (father)		
No father	0.025	0.157
Other inactive	0.009	0.096
Housework	0.002	0.047
Retired	0.010	0.101
Unemployed	0.008	0.091
Self-employed	0.181	0.385
Employed	0.763	0.425
Labour market status (mother)		
No mother	0.008	0.087
Other inactive	0.011	0.105
Housework	0.364	0.481
Retired	0.005	0.068
Unemployed	0.007	0.086
Self-employed	0.097	0.296
Employed	0.508	0.500
Main occupation (father)		
No father	0.155	0.361
Father not working	0.028	0.163
Armed forces occupation	0.010	0.099
Elementary occupations	0.089	0.285
Plant and machine operators and assemblers	0.120	0.325
Craft and related trades workers	0.209	0.407
Skilled agricultural, forestry and fish	0.099	0.299
Service and sales workers	0.055	0.228
Clerical support workers	0.038	0.191
Technicians and associate professionals	0.079	0.269
Professionals	0.068	0.252
Managers	0.050	0.218
Main occupation (mother)		
No mother	0.133	0.339
Mother not working	0.348	0.476
Armed forces occupation	0.001	0.024
Elementary occupations	0.097	0.296
Plant and machine operators and assemblers	0.030	0.171
Craft and related trades workers	0.044	0.204
Skilled agricultural, forestry and fish	0.066	0.249
Service and sales workers	0.093	0.290
Clerical support workers	0.065	0.246
Technicians and associate professionals	0.050	0.218
Professionals	0.059	0.236
Managers	0.015	0.121
Managerial position (father)	0.220	0.414
Managerial position (mother)	0.059	0.235

Country of birth (father)		
Present country	0.865	0.342
Other 27	0.072	0.258
Other Europe	0.021	0.145
Outside Europe	0.032	0.177
No parent, unknown birth	0.010	0.097
Country of birth (mother)		
Present country	0.872	0.334
Other 27	0.073	0.260
Other Europe	0.022	0.146
Outside Europe	0.033	0.178
No parent, unknown birth	0.001	0.024
Citizenship (father)		
Present country	0.891	0.311
Other 27	0.058	0.234
Other Europe	0.019	0.136
Outside Europe	0.028	0.164
No parent, unknown birth	0.004	0.062
Citizenship (mother)		
Present country	0.895	0.306
Other 27	0.057	0.232
Other Europe	0.019	0.135
Outside Europe	0.028	0.164
No parent, unknown birth	0.001	0.033
Tenancy status		
Tenant	0.231	0.421
Accommodation was provided free	0.049	0.216
Owner	0.720	0.449
Financial situation of the household		
Very bad	0.038	0.190
Bad	0.084	0.278
Moderately bad	0.167	0.373
Moderately good	0.398	0.490
Good	0.266	0.441
Very good	0.048	0.213

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

Table A.5

SUMMARY STATISTICS, 2019

	Mean/Prop.	Std. Dev.
Panel A: Individual and household characteristics		
Age	41,647	8,785
Female	0.512	0.500
Migrant status		
Native	0.883	0.321
Migrant from EU origin	0.042	0.200
Migrant from another region	0.075	0.264
Level of education		
Primary	0.176	0.381
Secondary	0.468	0.499
Tertiary	0.356	0.479
Labour market status		
Employed	0.802	0.398
Unemployed	0.078	0.268
Student	0.020	0.141
Other inactive	0.100	0.300
Panel B: Retrospective information on surveyed individuals		
Loneparent family	0.114	0.318
Number of children		
0	0.016	0.127
1	0.246	0.431
2	0.419	0.493
3	0.189	0.392
4	0.066	0.248
5 or more	0.063	0.242
Highest level of education (father)		
No father	0.036	0.186
Low	0.473	0.499
Medium	0.343	0.474
High	0.148	0.355
Highest level of education (mother)		
No mother	0.010	0.097
Low	0.526	0.499
Medium	0.345	0.475
High	0.120	0.325
Labour market status (father)		
No father	0.035	0.184
Other inactive	0.012	0.109
Housework	0.002	0.050
Retired	0.007	0.084
Unemployed	0.009	0.928
Self-employed	0.184	0.387
Employed	0.751	0.433

Labour market status (mother)		
No mother	0.009	0.096
Other inactive	0.022	0.147
Housework	0.318	0.466
Retired	0.004	0.062
Unemployed	0.010	0.099
Self-employed	0.096	0.295
Employed	0.540	0.498
Main occupation (father)		
No father	0.086	0.280
Father not working	0.069	0.254
Armed forces occupation	0.009	0.097
Elementary occupations	0.086	0.280
Plant and machine operators and assemblers	0.120	0.325
Craft and related trades workers	0.207	0.405
Skilled agricultural, forestry and fish	0.085	0.279
Service and sales workers	0.077	0.266
Clerical support workers	0.045	0.207
Technicians and associate professionals	0.086	0.280
Professionals	0.079	0.270
Managers	0.050	0.218
Main occupation (mother)		
No mother	0.029	0.168
Mother not working	0.379	0.485
Armed forces occupation	0.001	0.031
Elementary occupations	0.093	0.291
Plant and machine operators and assemblers	0.031	0.172
Craft and related trades workers	0.050	0.218
Skilled agricultural, forestry and fish	0.060	0.237
Service and sales workers	0.117	0.322
Clerical support workers	0.082	0.274
Technicians and associate professionals	0.059	0.235
Professionals	0.082	0.274
Managers	0.017	0.131
Managerial position (father)	0.151	0.358
Managerial position (mother)	0.054	0.226
Country of birth (father)		
Present country	0.775	0.417
Other 27	0.045	0.208
Other Europe	0.065	0.247
Outside Europe	0.072	0.258
No parent, unknown birth	0.043	0.201
Country of birth (mother)		
Present country	0.807	0.395
Other 27	0.048	0.214
Other Europe	0.068	0.252
Outside Europe	0.076	0.266

Citizenship (father)		
Present country	0.798	0.401
Other 27	0.041	0.199
Other Europe	0.060	0.238
Outside Europe	0.066	0.248
No parent, unknown birth	0.034	0.182
Citizenship (mother)		
Present country	0.820	0.384
Other 27	0.041	0.198
Other Europe	0.061	0.239
Outside Europe	0.069	0.253
No parent, unknown birth	0.009	0.095
Tenancy status		
Tenant	0.184	0.388
Accommodation was provided free	0.034	0.181
Owner	0.782	0.413
Financial situation of the household		
Very bad	0.026	0.159
Bad	0.068	0.251
Moderately bad	0.144	0.351
Moderately good	0.398	0.489
Good	0.307	0.461
Very good	0.057	0.232
Basic school needs	0.047	0.211
Daily protein	0.082	0.274
One week annual holiday away from home	0.278	0.448
Degree of urbanisation		
Rural area	0.473	0.500
Town or suburb	0.313	0.464
City	0.214	0.410

Source: Authors' computation, using data from the cross-sectional component of the European Union – Statistics on Income and Living Conditions (EU-SILC).

APPENDIX B

As explained above, our estimates of the upper bound of IOp are derived using the longitudinal component of the EU-SILC. We build our sample in a form that uses all the information available across data releases and not the one contained in a single data release.

There are two important aspects that one needs to consider when dealing with the longitudinal component of the EU-SILC. First, the EU-SILC is a rotational panel. In most countries, there are four rotational groups which are each interviewed for four consecutive waves. Each year, the households that have been interviewed for the fourth time leave the panel and are substituted by a new rotational group that is interviewed for the first time. As a result, 25% of the sample is renewed each year. A second additional issue is how data is released by Eurostat to the research community. In this respect, one needs to take into account that in order for a household to be included in a data release it needs to comply with two criteria: 1) the household is part of the sample for the latest wave of data collection, and 2) the household was in the sample the year before the data release (Iacovou and Lynn, 2013). This means that, for example, households that are interviewed in 2016 for the first time (rotational group 4 in Table B.1) will not be part of the 2016 data release. These households will appear for the first time in the 2017 data release when information relative both to 2016 and 2017 will be provided.

Table B.1

LONGITUDINAL PANEL STRUCTURE AND DATA RELEASES FOR SPAIN, EU-SILC, 2016–2020

	Rotational group 1	Rotational group 2	Rotational group 3	Rotational group 4
2016 release	2013 2014 2015 2016	2014 2015 2016	2015 2016	
2017 release		2014 2015 2016 2017	2015 2016 2017	2016 2017
2018 release	2017 2018		2015 2016 2017 2018	2016 2017 2018
2019 release	2017 2018 2019	2018 2019		2016 2017 2018 2019
2020 release	2017 2018 2019 2020	2018 2019 2020	2019 2020	

Note: Table adapted from Figure 1 in IACOVOU and Lynn (2013).

This panel structure has two important consequences for the construction of a working sample: a) if one simply takes a data release for a given year, s/he will be missing 25% of the sample as one rotational group is always missing in every data release; and, b) simply combining the different releases means that households will be duplicated multiple times. In order to deal with both issues, and as recommended by Iacovou and Lynn (2013), we build our working sample by taking all household/year information from the last wave in which a rotational group participated – note the coloured cells in Table B.1 – except for the last year of data for which we take the information available for all rotational groups. This way we ensure that data for each year contains 100% of the sample – note for example data for 2016 in bold –, that our sample does not suffer from duplicated observations, that data covers the longest run of data available for each rotational group and that variable changes between two waves for the same rotational panel do not affect our sample.