Opportunity bias in Spain: Empirical evidence, drivers and trends

Marta Curull Sentís

Supervised by:
Sara de la Rica and Lucía Gorjón (ISEAK)

2020/21
Master in Economics: Empirical Applications and Policies
University of the Basque Country UPV/EHU

Acknowledgments

I would like to thank the ISEAK Foundation for giving me the opportunity to work with them in this study. The results presented in this research are filled with the knowledge and guidance of Sara and Lucía who have inspired and helped in every step of the way. I also acknowledge Antonio Villar for making their research and advice available for this paper. A special thanks to the EAP master faculty for their suggestions and ideas during this Master Thesis Defense.

The aim of this study is to measure opportunity bias in Spain. This purpose requires to analyse current outcomes as a function of opportunities, and how the latter affect the first. Opportunities are defined in terms of those circumstances of the individuals that are out of their control. It is crucial that these characteristics are exogenous, which we ensure by using information of the individual and her household when she was around 14 years old. The variation in educational opportunities is determined by circumstances in an 80%, mainly due to family origin. The share is between 30 and 45% for labour situation and hourly wage. Labour market gaps are partly shaped by family origin, but regional differences have the largest impact. In 2019, more than half income inequality is generated by inequality of opportunity, a share that has steadily increased since 2004.

Keywords: Inequality, inequality of opportunity, social justice, labour market, education.

1. Introduction

Spain has one of the largest levels of inequality in Europe. In 2019, it was higher than almost any other EU country, including the Mediterranean ones, the only exceptions were Bulgaria, Lithuania, Latvia and Romania, according to Eurostat[1]. The Gini coefficient increased between 2002 and 2016. Since 2017 it decreased, but inequality has not yet diminished enough to reach the levels prior to the Great recession, as it can be seen in *Figure 1*. In addition, it can be expected that gaps will increase again due to the current economic crisis generated after the outbreak of COVID-19. In this setting, our research focuses on a specific part of the disparities, the ones caused by characteristics of the individual that are out of their control. We analyse Spanish inequality of opportunity in 2019 by quantifying the impact and relevance of several circumstances on a set of outcomes related to the individuals' labour market performance.

Fig.1 **Evolution of inequality in Spain and Europe (Gini)**

Data source: Eurostat (left) and Observatorio de la pobreza y desigualdad en España, ISEAK (right)

Recent literature has studied inequality as being the result of two components, one related to effort and autonomous decisions of the individual and the other determined by circumstances out of their control (Roemer, 1993 and 1998; Van de gaer, 1993). This last component is inequality of opportunity (IO), which is the fraction of inequality that offsets social justice: if inequality of opportunity dominates, outcomes are less linked to skills, merit or effort and are a result of origin and family background. Additionally, although traditional research does not agree on whether the effects of inequality are good or bad for economic performance, newer investigations have found that the opportunities component is clearly negative for economic growth and efficiency (Marrero and Rodríguez, 2013 and 2019; Bradbury and Triest, 2016).

Spain is found to be one of the European countries with the highest levels of IO (Palomino *et al.*, 2019; Marrero and Rodríguez, 2012; Rodríguez, 2008). Furthermore, inequality of opportunity is increasing in this country. There is empirical evidence that social mobility has diminished in recent years, particularly for some demographic groups (Esping-Andersen *et al.*,

[1] https://ec.europa.eu/eurostat/databrowser/view/ILC_DI12__custom_1138823/default/table

2020; Cantó and Ruiz, 2015). According to an OECD report (2018)[2], income mobility in Spain is low both at the top and bottom parts of the distribution, and it has been shrinking since the 1990s.

Being able to identify and quantify inequality of opportunity and its drivers allows for more efficient public policies aimed at fostering social mobility. High levels of intergeneration transmission are related to lower levels of participation, life satisfaction and well-being and government trust (OECD, 2018). In addition, measuring IO is key to policies aimed at fostering social justice, as circumstance-based inequality is undoubtedly seen as unfair. In addition, if IO was the dominant component of overall inequality, its levels would have a negative impact on economic performance.

The aim of this study is to contribute into the literature of IO both by using an innovative methodology, created by Herrero and Villar (2018 and 2020), and by providing a dynamic study of opportunity bias importance using the most recent available data in Spain. We analyse current outcomes as a function of exogenous opportunities, and quantify how the latter affect the first. We start by measuring IO in a set of related labour market outcomes, and then, we compute its relevance and evolution in overall income inequality.

Regarding our first contribution, we use the opportunity advantage index developed by Herrero and Villar to compare distributions in terms of relative advantages. By means of this procedure, we quantify the impact of circumstances on three outcomes that are related to the labour market and essential for individual chances in life. We find that 80% of the variation in educational opportunities is determined by circumstances, this share is between 30 and 45% for labour situation and hourly wage. Labour market gaps are partly shaped by family origin, but regional differences are also large and significant.

For the second, we measure the importance of IO in Spain in 2019, the most recent data, and we compare it to 2011. We find that circumstances of the individual when she was young determine more than 50% of their current level of income. This share has steadily increased since 2004, as we can conclude by comparing our results with the study of Marrero and Rodríguez (2012). Household origin and background are the main source of this kind of inequality.

This paper is structured as follows. First, we describe and present the data and the set of circumstances used in the study. Then, we split the paper into two parts. Section 3 is devoted to the drivers of IO on their impact on three outcomes. We introduce the methodology used, describe the outcomes and present the main results. The second part of the analysis is presented in Section 4, where we measure the relevance of IO on overall inequality. This section includes its own methodology and the second set of results. Finally, we summarize our main findings and conclude.

[2] OECD (2018), *A Broken Social Elevator? How to Promote Social Mobility*, OECD Publishing, Paris, https://doi.org/10.1787/9789264301085-en

2. Data

We use individual microdata from the 2011 and 2019 Survey on Income and Living Conditions (SILC) of the Spanish Statistical Institute. This is a yearly database that complies data on income and living conditions by means of household interviews, combined with income administrative data. Both waves include a specific module devoted to analyse intergenerational poverty transmission with information about the conditions of the individuals and the characteristics of their households when they were around 14 years old. This specific module is included in the SILC every six to eight years. Using it, we are able to generate a rich dataset including data on individual's exogenous circumstances, which allows us to precisely estimate inequality of opportunity in Spain.

Our study analyses nowadays IO. Therefore, we use 2019 data throughout all the research. After presenting current results, we incorporate information from 2011 to measure the evolution of inequality of opportunity during this period.

Focusing on 2019, the original database included 39,852 observations. For our analysis we select individuals between 25 and 59 years old[3], for whom the SILC includes information on the background circumstances, resulting in a sample of 17,615 observations from 10,459 households.

Our first goal is to understand how much of the labour market outcomes of an individual are determined by their circumstances, i.e., characteristics out of their control. For the first part of our study, these will be gender, the region in Spain where the individual lives[4] and the family economic background when she was 14 years old. This last variable is discrete and includes six categories regarding the economic situation of the household when the individual was a teenager, ranking from very bad to very good[5]. The distribution of these circumstances among the Spanish population according to our data can be found in *Table 1*.

Table 1. Circumstances used for measuring the divers of IO

		N	Frequency
Gender	**Men**	11,370,081	50.02%
	Women	11,362,129	49.98%
Region	**Andalusia**	4,149,569	18.25%
	Aragón	617,782	2.72%
	Asturias	475,675	2.09%
	Canary Islands	1,154,364	5.08%
	Cantabria	274,560	1.21%
	Castilla León	1,084,685	4.77%
	Castilla Mancha	980,260	4.31%
	Catalonia	3,657,682	16.09%
	Extremadura	504,277	2.22%
	Galicia	1,273,962	5.60%
	Balearic Islands	629,320	2.77%

[3] We have also discarded observations from Ceuta and Melilla, the two Spanish autonomous cities, as the number of observations for these regions is very small.

[4] The SILC database does not include information on the region of birth. A robustness check for this issue can be found on Annex A using the Spanish Labour Force Survey.

[5] The information in this variable is the answer to the question "What was your economic situation when you were around 14 years old?". The answer to this question might be subjective. In further steps of this research, we will repeat, as a robustness check, the same analysis with other variables that are more objective.

Table 1. Circumstances used for measuring the divers of IO

	Rioja	148,858	0.65%
	Madrid	3,319,913	14.60%
	Murcia	738,506	3.25%
	Navarre	308,965	1.36%
	Basque Country	1,005,054	4.42%
	C. Valenciana	2,408,778	10.60%
Family Background	**Very Bad**	500,246	2.20%
	Bad	1,797,331	7.91%
	Moderately Bad	3,552,661	15.63%
	Moderately Good	8,882,694	39.08%
	Good	6,811,721	29.97%
	Very Good	598,012	2.63%
Total		**22,732,210**	

We link these circumstances with three relevant outcomes for the individuals' performance in the labour market. These are the highest attained level of studies, labour situation and hourly wage. All are presented as categorical variables in four-level intervals. For studies we use Less than secondary, Secondary, Post-secondary and Superior. For labour situation: Long term unemployment (more than 12 months), Short term unemployment (up to 12 months), Temporary contract and Indefinite contract. Finally, hourly wage is a continuous variable that we transform into categorical, so it is comparable with the others. We do so by setting the thresholds at the sample mean (11.3€/hour) and at the mean of each of the population subgroups obtained from dividing the society into those above and below the mean (18.8€/hour and 6.6€/hour, respectively). Using these three values we obtain four categories for the hourly wage variable. This procedure minimizes the error introduced when substituting a continuous distribution by a discrete variable with a limited number of intervals, as proposed by Esteban, Gradín and Ray (2006).

After quantifying and understanding the drivers of opportunity biases in the three mentioned outcomes, the second objective of this study is to measure what part of overall inequality is due to inequality of opportunity (IO). For this part, our variable of interest is household income, which includes all sources of income, adjusted by the number of household members using the OECD-modified equivalent scale. In this section, due to its methodology, we are able to increase the detail of family background circumstances by including a wide range of relevant and exogenous variables related to the situation of the individual and her household when she was young. With this dataset we are able to characterize in high detail individuals' origin and background. The description of the extra[6] circumstances can be found in the next table.

Table 2. Circumstances used for quantifying the relevance of IO on total inequality

		N	Frequency
Parental education	**Secondary or less**	15,766,553	69.36%
	Post-secondary	2,792,235	12.28%
	Superior	3,368,037	14.82%
Parental occupation	**Not working**	837,017	3.68%

[6] We include extra variables because the procedure used to estimate the importance of IO on total inequality allows for a lower number of observations in each cell generated by crossing all the circumstances. In order to have more individuals per cell, in the first part we summarize family background information in a unique variable. See the methodological sections 3.1 and 4.1 for more detail.

Table 2. Circumstances used for quantifying the relevance of IO on total inequality

	Low skilled	2,479,912	10.91%
	Middle skilled	14,341,340	63.09%
	High skilled	4,312,795	18.97%
Parental ethnicity	**At least one foreign-born**	2885741	12.69%
	At least one foreign-born (EU)	919403	4.04%
	All Spanish	18,213,195	80.12%
Type of family	**Lived with both parents**	19,968,136	87.84%
	Otherwise	2,764,074	12.16%
Number of siblings	**None**	4,912,830	21.61%
	One or two siblings	13,479,329	59.30%
	More than 3	3,661,557	16.11%
Did the mother work?	**No**	14,364,957	63.19%
	Yes	8,367,253	36.81%
Number of inhabitants	**Large num. inhabitants**	8,243,592	36.26%
	Med num. inhabitants	8,025,974	35.31%
	Small num. inhabitants	5,860,068	25.78%
Total		**22,732,210**	

3. The drivers of inequality of opportunity

The first goal of our analysis is to understand the drivers of inequality of opportunity. This is done by using an innovative methodology, developed by Herrero and Villar (2020), devoted to analyse opportunity advantages. One of the contributions of their approach is that it offers a relative and precise quantitative evaluation of the advantages and disadvantages of population subgroups in terms of opportunities. We apply this method to a diverse set of outcomes related to individuals' labour market performance.

The most well-known definition of inequality of opportunity among economists was proposed by Roemer (1993; 1998) who defined outcomes as the result of two components, *effort* and *opportunity*. According to that definition, the first one includes all that refers to agents' external circumstances, given at birth or at a young age, and thus, out of their control. The second is related to responsibility and autonomous choices of individuals[7].

Roemer's characterisation considers that there is equality of opportunity if all individuals who exert the same degree of effort obtain the same level of outcome. In other words, same effort should lead to the same results, independently of individuals' origin. This is known as the *ex-post* approach. Van de gaer (1993) proposed an alternative view: there is equality of opportunity if the opportunity set available to all individuals is independent of their initial circumstances. This is the *ex-ante* approach, the one we adopt in this study. Following Van de gaer theorization, by dividing the population into subgroups defined by exogenous circumstances, we can measure the impact of inequality of opportunity as the between-group differences in each of the studied outcomes.

[7] Nevertheless, *effort* is very difficult to measure and can be highly related to circumstances. In addition, other unobservables, like luck, also play a role and can influence *effort*. This is why we focus our study on analysing the impact of circumstances, and we don't try to measure or quantify *effort*. Specifically, we study the overall effect of circumstances which includes the direct and indirect effect (the impact of circumstances through "effort" and other unobservables, like luck, etc.).

We focus on three outcomes, education levels, job status, in terms of type of contract and duration of unemployment, and hourly wage. Empirical evidence finds a strong relation between the level of studies and labour market outcomes. Measuring the importance of circumstances in these three variables provides an idea of the impact of inequality of opportunity on labour market differences.

Previous literature concludes that circumstances play an important role in Spain. Choi *et al.* (2017) find that inequalities in educational attainment generated by socioeconomic circumstances, gender and place of birth originate at lower educational levels in Spain. Esping-Andersen *et al.* (2020) analyse whether the social class of individuals, measured in terms of occupational positions, is more influenced by their skills or by their family background. They compare several countries and find that Spain and Italy are the two places where the impact of social origin in determining access to a high social position is the largest. Family origin and social environment have also gained relevance in the allocation of people's chances in life, reducing social mobility. A fact that is specially reflected in the social structure of unemployment and the distribution of its social costs (Gorjón, de la Rica and Villar, 2020).

Moreover, Herrero, Villar and Soler (2018) apply the opportunity advantage index on Spanish income distributions before, during and after the 2008 economic crisis. Despite finding large regional differences, they conclude that overall, in 2016, Spain had 15% less income opportunities than before the recession and the reduction on employment opportunities was of 20%.

By analysing our three labour market outcomes we contribute to IO research in Spain, particularly by applying a novel procedure, the opportunity advantage index. In the next section we present this innovative methodology, and after that, our results.

3.1 Methodology

Herrero and Villar (2018; 2020) approach offers a formal relative measure to rank and compare, in a complete and cardinal way, outcome distributions in terms of opportunity advantages. Specifically, it divides the population into subgroups and measures the probability of a random individual in a subgroup to obtain higher outcomes than another random individual in a different subgroup. The method results in a precise indicator that attaches a value to each distribution that is proportional to the likelihood of getting higher outcomes.

As mentioned, the key element that allows this indicator to be used as a measure of opportunity inequality is the fact that subgroups are defined in terms of circumstances, which are, by definition, out of the control of the individual. Exogeneity is essential as it allows us to quantify the extent to which current outcomes are determined by characteristics of individuals that, in a setting of equality of opportunity, would be independent of the outcomes.

We use the opportunity advantage index in its discrete form, as the three labour market outcomes that we study are categorical. In this way, we obtain a number for each subgroup that indicates how many opportunities the subgroup has, relative to a reference group. This valuation is calculated as follows:

$$\varphi_i = \frac{\sum_{j \neq i} q_{ij} \varphi_j}{\sum_{j \neq i} q_{ji}} \tag{1}$$

Where φ_i represents the valuation of subpopulation i, defined by exogenous circumstances. Then, the numerator represents the overall advantage of i over j, q_{ij}, adjusted by the valuation of subpopulation j, φ_j. The denominator is a proportionality degree equal to the overall disadvantage of i over j. Overall advantages and disadvantages are calculated comparing i and j distributions:

$$p_{ij} = a_{iz}\left(a_{j(z-1)} + \cdots + a_{j0}\right) + a_{i(z-1)}\left(a_{j(z-2)} + \cdots + a_{j0}\right) + \cdots + a_{i1}a_{j0}$$

$$q_{ij} = p_{ij} + \frac{e_{ij}}{2}$$

In the first formula a_{iz} represents the share of individuals in subpopulation i with level z of the outcome we evaluate. For example, if we divide the population into binary gender subgroups, women (i) and men (j), and we want to evaluate the level of studies measured in the four mentioned categories, z would be equal to 4 and a_{iz} would be the proportion of women having superior studies and $a_{i(z-1)}$ would be the proportion of women having post-secondary education. Similarly, $a_{j(z-1)}$ would be the proportion of men having post-secondary education and a_{j0} would be the proportion of men having less than secondary education. This way, in the expression above, the overall advantage of women over men, q_{ij}, is obtained by calculating how often a woman obtains a higher level of studies than a man, plus half of the probability that a woman and a man obtain the same level of studies, $\frac{e_{ij}}{2}$[8].

The summatory in equation (1) is there to indicate comparison between more than two subgroups. When there are multiple subpopulations, the procedure is repeated to all pair-wise comparisons and solved altogether, using consistency and proportionality requirements. See Herrero and Villar (2020) for more detail[9].

We first apply Herrero and Villar's index into the three sets of circumstances: gender (2), region (17) and family background (6) and present a description of subgroup advantages and disadvantages. Once we cross these characteristics, we obtain 204 cells (2x17x6)[10]. Then, we attach a valuation based on the indicator to each of these cells for the three outcomes of interest: level of studies, labour situation and hourly wage. Doing a cell-based analysis requires the number of cells to be limited, in order to have enough information per cell. This is the reason why we use a unique variable summarizing family background, instead of using a larger and more detailed set of variables.

The purpose of the construction of this cells-dataset is to be able to estimate the relevance of each circumstance on the index. We run OLS regressions where the dependent variable is each of the three indexes, measured in these 204 cells, and the independent variables are the

[8] The probability of an individual in i obtaining the same outcome level as an individual in j is the same for both subgroups: $e_{ij} = e_{ji}$. As $p_{ij} + p_{ji} + e_{ij} = 1$, we can calculate $q_{ij} = p_{ij} + \frac{e_{ij}}{2}$, and thus, $q_{ij} + q_{ji} = 1$.

[9] An algorithm for the computation of the Opportunity advantage index is freely provided by Ivie. https://web2011.ivie.es/balanced-worth/

[10] Two cells are empty for level of studies and labour situation and four cells are empty for hourly wage.

circumstances. This analysis allows us to check whether there is a relation between circumstances and opportunity biases and which associations are the most important. We also want to see whether the opportunity bias in the level of studies affects the labour situation outcome, and if these two have an effect on hourly wage. That's why we run extra regressions including the previous[11] indexes.

The OLS regressions we use in this setting are summarized in the next formula, where the subindex i takes three values, one for each outcome, and the subindex k relates to the 204 cells.

$$opportunity\ bias_{ik} = \beta_0 + \beta_1 gender_k + \beta_2 region_k + \beta_3 family\ background_k + \varepsilon_k$$

In addition, we also study the effect of the circumstances and previous opportunity differences on the outcomes per se. This is why we repeat the previous regressions, but now the dependent variables are the three outcomes. As we are still using the cell-database, we transform the outcomes in order to obtain a representative value per cell. For the level of studies, we use the share of individuals in each cell with superior studies, for labour situation the proportion of employed and for hourly wage we use the mean of the cell. In this case, the subindex j takes three values, one per each outcome, and the subindex k refers to the cells.

$$outcome_{jk} = \beta_0 + \beta_1 gender_k + \beta_2 region_k + \beta_3 family\ background_k + \varepsilon_k$$

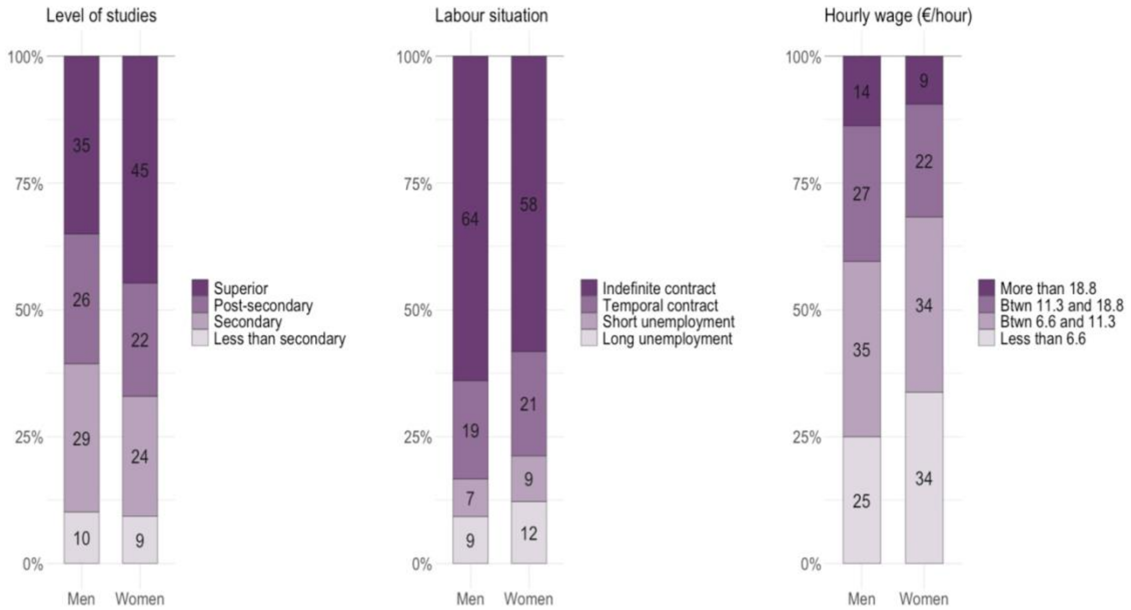
3.2 The relation between outcomes and circumstances – Descriptive findings

In this section we study the distribution of each outcome by circumstances. It is important to underline that this research links current outcomes with given characteristics, out of the individuals' control.

Figure 2 depicts the distribution of outcomes by gender. Women obtain better results in education, as they present higher percentages in high level of studies and lower ratios in low levels. However, this does not translate into a better labour situation: there are more women in long and short unemployment and more women with temporary contracts, compared to men; females have also lower salaries.

[11] Previous in terms of the moment these outcomes are acquired. It is understood that higher attained level of studies lead, to a certain degree, to better job statuses and higher salaries. In turn, improving once's job status is usually associated with larger wages.

Fig.2 Outcome distribution by gender (in percentages)



There are substantial differences by region (*Figure 3*). For instance, Madrid and the Basque Country have the highest rate of superior studies (53%), whereas Murcia has less than half this percentage (25%). The percentages for the top performance in labour situation, which corresponds to having an indefinite contract, are larger than in the other categories. Despite that, we also observe large changes. Extremadura and Andalusia have around 40% of their workforce with indefinite contracts, while for most of the Spanish regions this rate is higher than 60%, and five autonomous communities reach, and some surpass, 70%. Regarding hourly wage, we can see that in Extremadura half of the population earns less than 6.6 euros per hour. On the other extreme there is the Basque Country, where 19% have an hourly salary above 18.8 and more than half its inhabitants earn more than 11.3 euros per hour; this fact is only reproduced in Navarre.

Fig.3 Outcome distribution by region (in percentages)

In *Figure 4* we plot current outcome distributions per family background, the economic situation of the household when the individual was a teenager. In this case we observe a clear relation between the circumstance and the performance in the labour market. As the economic family background improves, the percentage of individuals with superior studies quadruplicates, from 14% to 63%. This is the largest increase, but the impact of a good economic situation when being a teenager also increases the ratio of individuals with indefinite contracts and higher salaries in a substantial way. Individuals with a good economic background have 30% more chances to get an indefinite contract than those with very bad backgrounds, and less than half the probability of being in long unemployment. The likelihood of a high salary more than doubles and that of a low hourly wage almost halves.

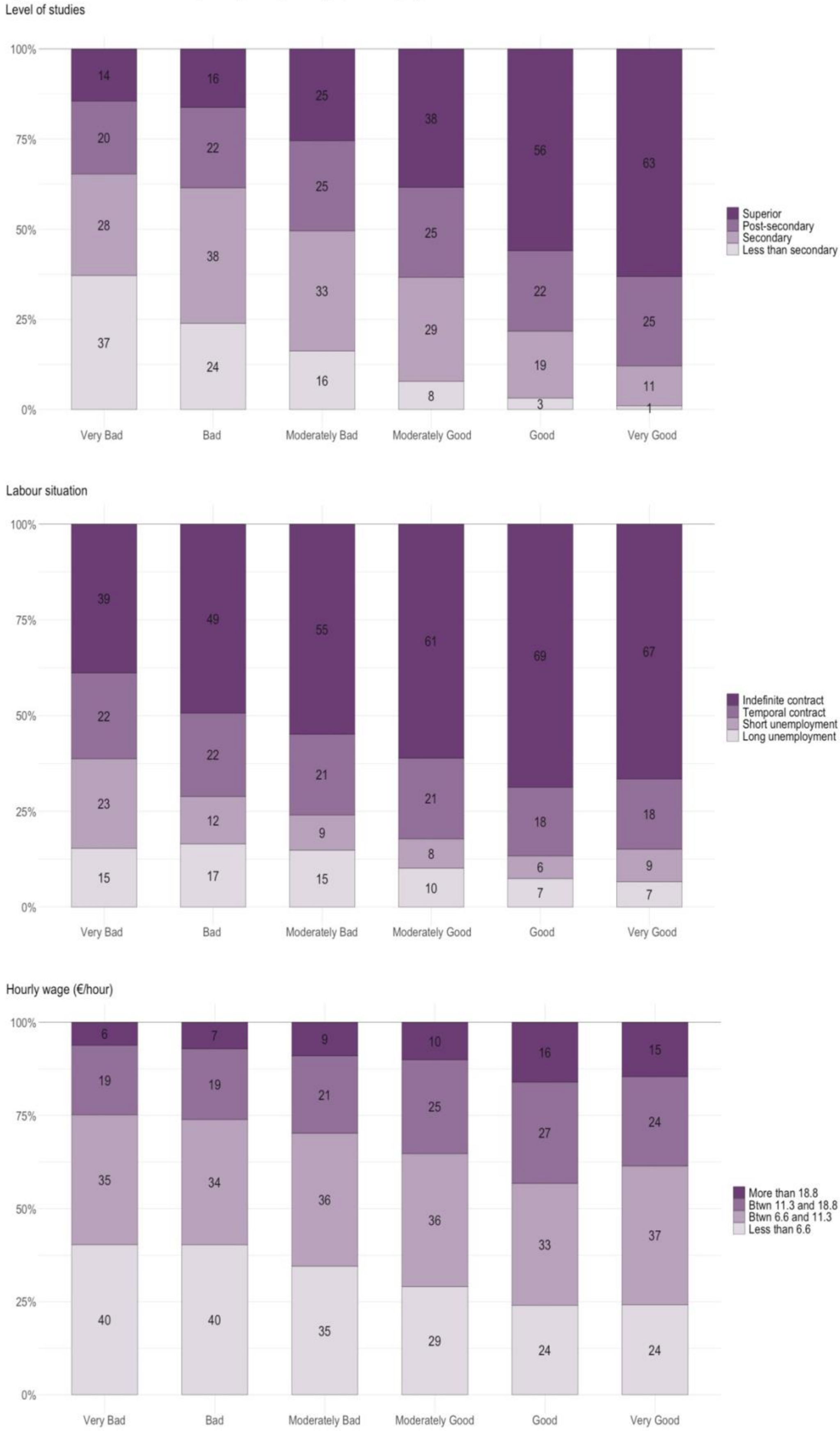
In this visual analysis, we have also crossed gender and family background to check whether the economic characteristics of the household where the individual grew up have different impacts depending on the individuals' gender. The results for gendered family background are presented in *Annex B*. In *Figure B1* we see that women obtain better educational levels than men in the same family background for all categories. For both genders, the chances of superior studies more than triple when family background is improved. This increase is especially relevant for women; 68% of those coming from the richest families obtain superior studies, the rate is 11 points lower for men.

In this top economic origin only 1% of men and women have less than secondary education, whereas women in the poorest families have a chance of 39% of having less than secondary education (35% is the share for men). It seems that among the poorest families, women have higher chances to attain superior studies, but also larger probabilities of not finishing secondary education.

The increased educational opportunities for women do not translate into better job positions: in almost all family background levels, men have a larger share of indefinite contracts, and women a larger presence in long unemployment. Only 1% of women with a very bad family origin reach a top position on the wage distribution, the share is 10% for men. In the wage allocation, men have larger opportunity advantages, but for both genders, those seem highly determined by family origin.

Overall, having the top category for salary is much less common than achieving the best result for the other outcomes. We find quite small percentages of individuals at the top of the distribution for hourly wage, whereas having an indefinite contract seems relatively more common in Spain and the probability of obtaining superior studies lies somewhere in between.

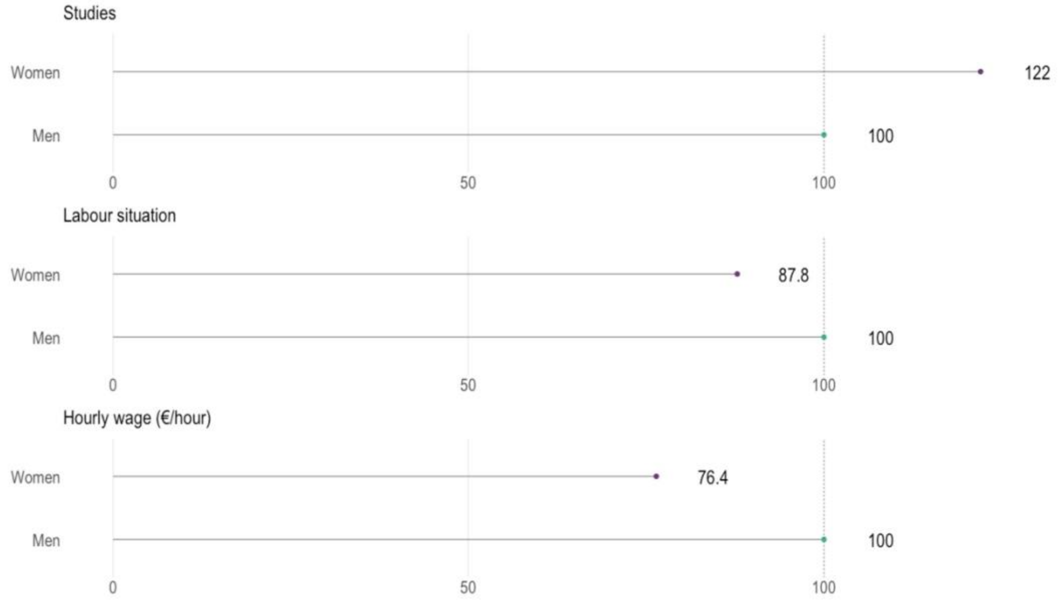
Fig.4 Outcome distribution by family background (in percentages)



3.3 The opportunity advantage index

After a descriptive analysis of the relationship between current outcomes and exogenous circumstances, we apply the opportunity advantage methodology (Herrero and Villar, 2020). This procedure reveals the same intuitions that we presented in the previous section, but offers a quantitative value for the magnitude of the opportunity bias of each subgroup, relative to a reference. In the following figure, we plot the opportunity bias by gender. We set men at one hundred and obtain the advantage or disadvantage of women with respect to men for each outcome. We can see that while women have a 22% advantage in studies, they have a 12.2% disadvantage in labour situation and a 23.6% disadvantage in hourly wage.

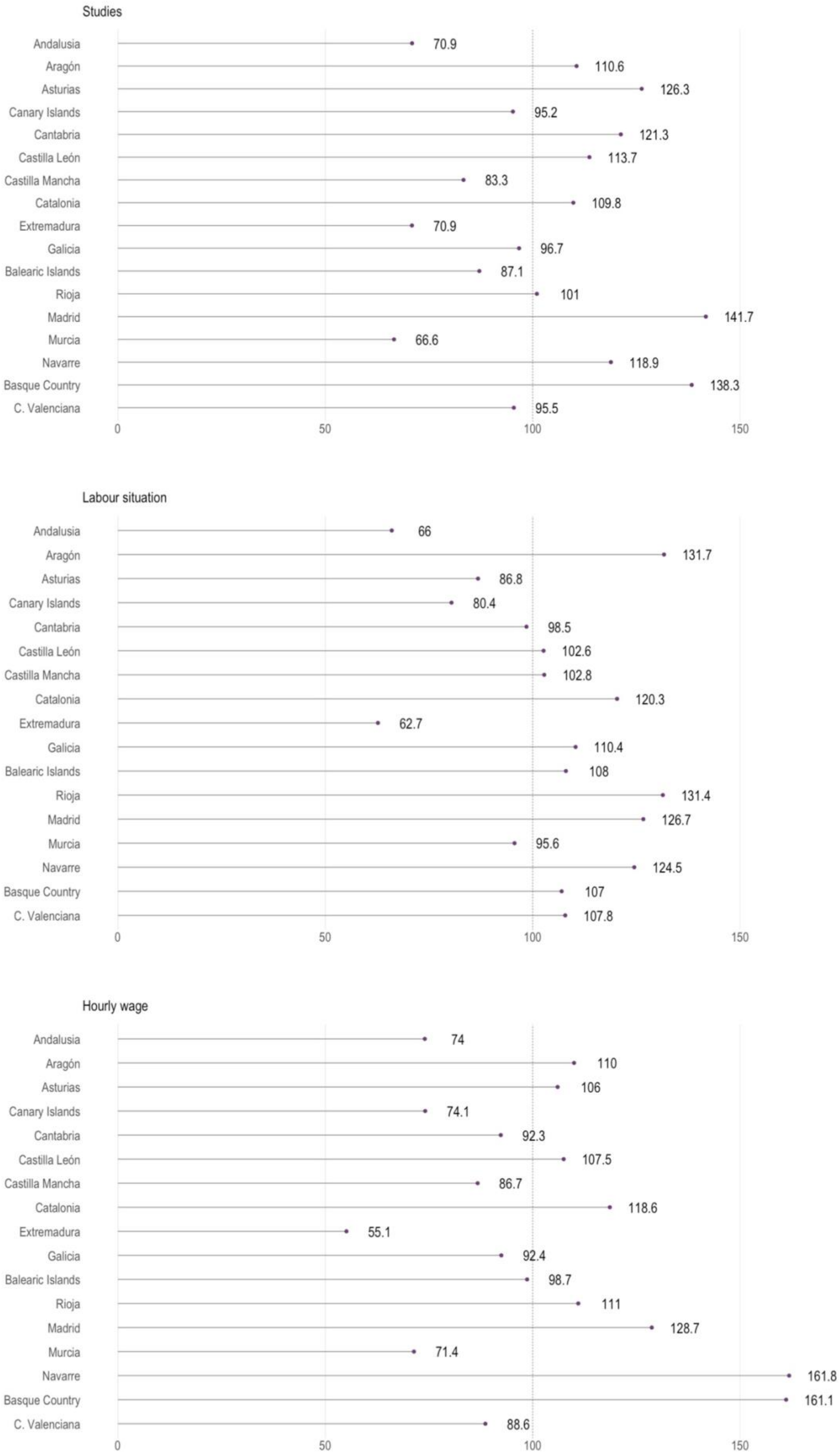
Fig.5 Opportunity bias by gender



The results for the opportunity advantage per region can be found in *Figure 6*. In this case, we have set Spain as a hundred, so the indicator refers to the opportunity of each region with respect to the whole country. In doing so, we can compare each region with the Spanish average and check whether it offers opportunities above or below the country's mean. For instance, the regions that bring the most opportunities in studies are Madrid, followed by the Basque Country; they provide an advantage around 40%, compared to Spain as a whole. On the other hand, the opportunities for education in Murcia are the lowest, with a disadvantage of more than 30%. This implies that in Murcia the likelihood of obtaining a high educational outcome is 30% lower than the Spanish average.

For labour situation, the top regions are Aragón and Rioja, with an advantage slightly higher than 30%. For this outcome, regions appear to be closer to the country's average, but two of them present a clear disadvantage of more than 30%, those are Extremadura and Andalusia. Finally, regarding hourly wage, two regions stand out, the Basque Country and Navarre, which provide to their citizens an advantage of more than 60%, implying that the inhabitants in these regions have a 60% chance to obtain higher wages than the Spanish average. On the other part of the opportunities, we find Extremadura, with a disadvantage of almost 45%.

Fig.6 Opportunity bias by region



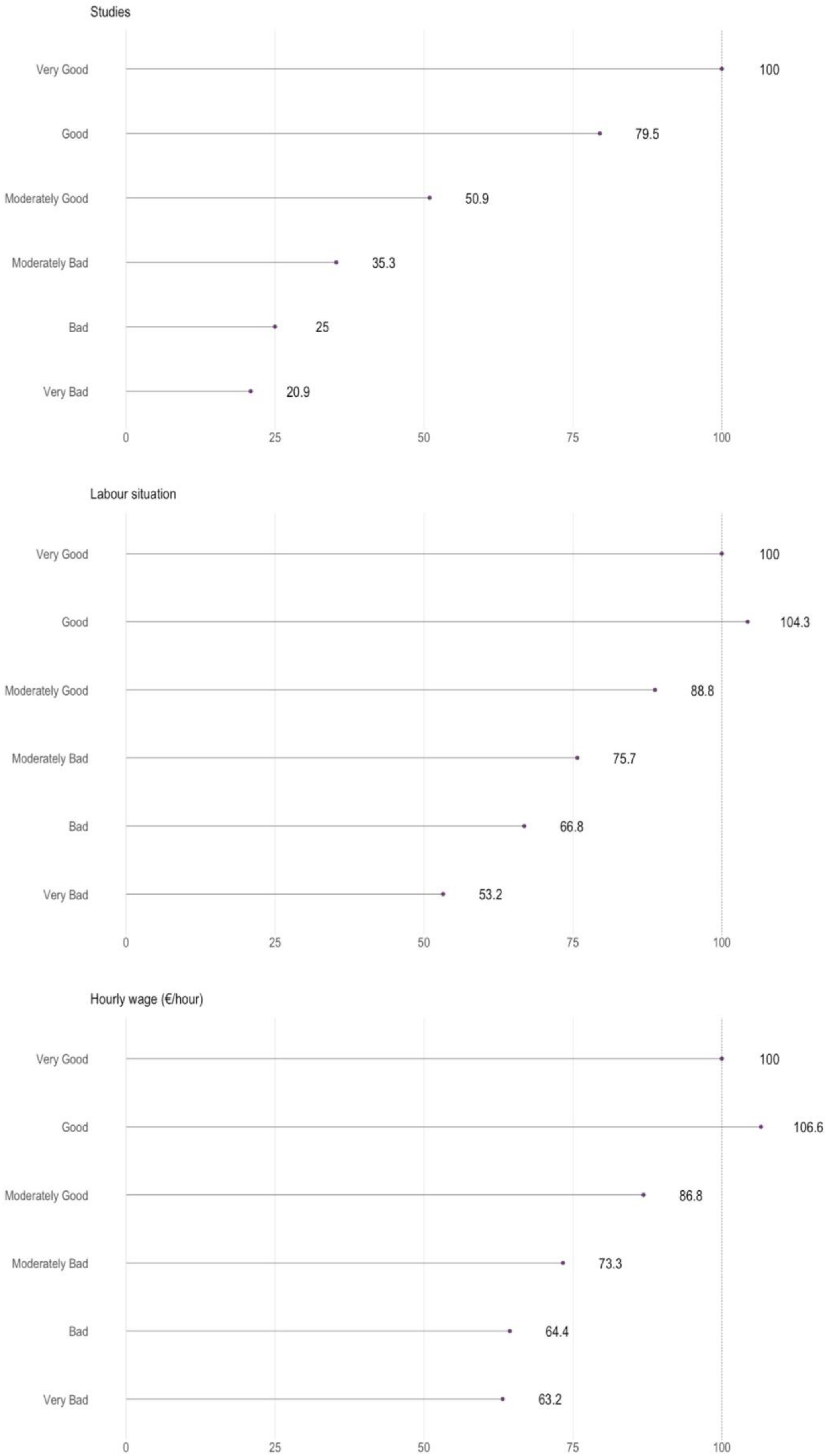
As mentioned before, SILC does not include information for the region of birth, only for the region where the individual is registered at the time of the interview. This might not be an exogenous circumstance as individuals may have moved from one region to another in order to achieve better outcomes. In Annex A we repeat the analysis using data from the Spanish Labour Force Survey, which includes both region of birth and registered region, but does not include wage variables, so the study is only redone for the level of studies and the labour status outcomes. This way we are able to compare the results and check whether there are substantial differences regarding the use of region of birth, as circumstance, or actual region. The correlation between region of birth and registered region is 0.76, which means that around one quarter of the individuals move. This generates some differences in the magnitude of the opportunity bias, mainly for the studies outcome. However, the ranking of opportunities is robust; the correlation between the studies index for region of birth and registered region is 0.95, and for labour situation is even higher, 0.97. That's why, from now on, we continue using region as almost an exogenous characteristic.

In *Figure 7* we plot the indexes by family background. We have set as the reference category those who had a very good economic situation when young. We uncover a clear relation between the opportunities an individual has at present and their economic origin. For all outcomes, the lower the economic security of the individuals when teenagers, the higher the disadvantage they face. There is only one exception, those who had a good economic situation seem to slightly outperform those with a very good economic situation in both the labour situation and hourly wage; their advantage is around 5%.

In sum, these results indicate that exogenous circumstances determine individuals' outcomes in life. For instance, we can see that those with very bad economic situations when young have a disadvantage of almost 80% in education. The disadvantage is almost 50% for labour situation and almost 40% for hourly wage.

These differences have also a gender component, as plotted in *Figure B2*. Now the reference group is men in a very good economic situation. Again, we see that women present relative advantages for education, but not for the other outcomes. For example, it seems that women from the richest origins have less chances to high labour outcomes both in their job status and hourly wage, but present a 22% advantage in education. What is even more concerning is the disadvantage for women from bad and very bad origins, they face a wage disadvantage around 60%.

Fig.7 Opportunity bias by family background



3.4 Estimating the drivers of the opportunity advantage index – Econometric results

In the last section of this first part, we estimate the outcomes of the cells defined above on all three circumstances in order to understand the extent to which different circumstances affect the outcomes under analysis. As mentioned, we also include previous indexes to control for accumulated opportunity biases.

The relation between opportunity advantages in one outcome and the others can be seized by computing the correlation between the indexes. It seems like having a high opportunity on one outcome does not necessarily lead to better opportunities in another. Labour situation is the most interrelated outcome: the opportunity index for labour status is linked with the opportunities in studies with a correlation of 0.28 and with the chances of hourly wage by 0.29. The linkage between opportunities in education and hourly wage is weaker, with a correlation of 0.2. This implies that, in a cell level, we can have subgroups with relative disadvantage one outcome but with relative advantage on another.

In Table 3 we show how each circumstance determines opportunity biases. The first relevant conclusion is found in the R-squared. Circumstances explain around 70% of the variance of opportunities in education, but around 30% of the variance of opportunities in labour situation and hourly wage.

In the opportunities of education (1), the most relevant factor is the household economic situation when the individual was young, which is reasonable, as education is mostly received while being young, therefore it is plausible that it is more affected by circumstances. For labour status opportunities (2) the most relevant circumstance seems to be the region: most regions increase labour opportunities compared to Andalusia, which is intuitive as, overall, this region has one of the lowest indexes for labour opportunities as seen in *Figure 6*. IO in studies explains only 8% of the variation in labour advantages (3), and it turns non-significantly different from zero when the circumstances are incorporated (4). Finally, for prospects in hourly wage (5), living in Navarre or in the Basque Country is positive and significant, but the next most important factor is being a women, which is related to a decrease in salary opportunities. Intuitively, good chances in labour status increase hourly wage opportunities (6), together with the education index they determine 10% of salary prospects. The labour indicator remains significant when adding the circumstances (7).

Table 3. The impact of the circumstances on each index

		Studies	Labour situation			Hourly wage		
Index		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Index - studies				0.347***	0.362		0.0616	-0.0507
				(0.102)	(0.282)		(0.0399)	(0.0987)
Index – labour situation							0.106***	0.0650**
							(0.0251)	(0.0296)
Gender *(base: Men)*								
	Women	4.133***	-9.085***		-10.58***	-5.357***		-4.549***
		(1.428)	(2.785)		(3.411)	(1.077)		(1.026)
Region *(base: Andalucía)*								
	Aragón	10.34**	32.96***		29.21***	1.628		0.00866
		(5.232)	(10.05)		(9.757)	(1.889)		(2.214)
	Asturias	8.489***	9.717***		6.642	4.347**		4.145**
		(2.872)	(3.233)		(4.096)	(1.803)		(2.001)

Canary Islands	2.751*	12.56**		11.56**	1.697		1.019
	(1.563)	(5.628)		(5.612)	(2.064)		(2.073)
Cantabria	13.61***	22.30***		17.37**	12.35		11.44
	(3.917)	(7.851)		(7.753)	(9.421)		(10.71)
Castilla León	7.013**	22.57***		20.03***	3.989*		2.876
	(3.183)	(3.879)		(4.692)	(2.153)		(2.237)
Castilla Mancha	3.138*	23.26***		22.12***	1.257		-0.0960
	(1.695)	(6.111)		(6.050)	(1.413)		(1.607)
Catalonia	9.573***	23.55***		20.08***	4.554***		3.508*
	(2.142)	(3.904)		(4.769)	(1.224)		(1.799)
Extremadura	6.510	4.761		2.403	1.613		1.634
	(5.010)	(6.293)		(4.854)	(2.300)		(2.119)
Galicia	10.26**	29.39***		25.67***	1.778		0.387
	(4.616)	(7.575)		(8.234)	(1.293)		(1.944)
Balearic Islands	2.887	25.12***		24.07***	5.890**		4.403*
	(2.157)	(5.638)		(5.348)	(2.455)		(2.613)
Rioja	6.680***	22.68***		20.26**	2.747		1.611
	(2.481)	(7.497)		(7.912)	(2.245)		(2.627)
Madrid	13.01***	27.91***		23.20***	3.758**		2.602
	(1.919)	(3.873)		(5.501)	(1.597)		(2.298)
Murcia	-3.175	15.16***		16.32***	-0.263		-1.411
	(2.941)	(3.759)		(4.180)	(1.392)		(1.468)
Navarre	6.414*	27.82***		25.50***	9.061***		7.395***
	(3.409)	(9.044)		(8.874)	(2.360)		(2.366)
Basque Country	14.54***	19.49***		14.22**	6.998***		6.467***
	(3.102)	(4.520)		(5.810)	(1.881)		(2.414)
C. Valenciana	1.836	26.58***		25.92***	1.232		-0.404
	(2.744)	(6.351)		(7.226)	(1.278)		(1.556)
Family background *(base: Very bad)*							
Bad	-0.247	-7.132		-7.042	-4.011**		-3.595**
	(1.740)	(6.358)		(6.209)	(1.550)		(1.555)
Moderately Bad	4.183**	-6.506		-8.021	-4.147**		-3.504**
	(1.614)	(5.854)		(5.978)	(1.641)		(1.623)
Moderately Good	11.17***	-0.493		-4.539	-2.693*		-2.087
	(1.607)	(5.751)		(6.590)	(1.618)		(1.790)
Good	21.86***	7.143		-0.774	-0.949		-0.298
	(1.605)	(5.697)		(8.480)	(1.596)		(2.465)
Very Good	39.42***	7.704		-6.576	3.306		4.746
	(3.767)	(7.254)		(12.23)	(2.779)		(5.824)
Constant	3.001*	29.41***	36.80***	28.33***	9.929***	2.977***	8.176***
	(1.758)	(5.315)	(2.866)	(5.284)	(1.779)	(1.142)	(2.068)
Observations	202	202	202	202	200	200	200
R-squared	0.716	0.276	0.080	0.301	0.293	0.101	0.312

Notes: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.
Models (1), (2) and (5) regress each index on the set of circumstances. Models (3) and (6) regress the indexes on the *previous* indexes: opportunities in labour situation as determined by opportunities in studies and opportunities in hourly wage, by opportunities in studies and labour situation. Models (4) and (7) use as regressors both circumstances and *previous* indexes.

Lastly, we check the impact of circumstances on the outcomes themselves. We repeat the previous estimations but the dependant variable, instead of being the opportunity advantage index is the outcome itself. For the studies outcome we use the share of individuals with superior studies in each cell, for labour situation the percentage of employed individuals in the cell and for hourly wage the average hourly wage per cell, this last regression is log-linear.

The results of these regressions can be found in Table 4. By using them, we can confirm that background characteristics are by far the most important circumstance for the level of studies but not so determinant for labour situation and hourly wage, though they are still relevant.

Almost 80% of the variation in the frequency of individuals that attain superior studies in each cell can be explained by circumstances (1). Being a woman increases this share and some regions also offer higher probabilities, but the most important factor is economic family situation. The top group increases the ratio of superior studies by almost 53 points. This last variable is also important in increasing hourly wage, coming from a very good economic background increases mean salaries for the cell by almost 43%, when controlling for opportunity bias in studies and job status (5). However, family background does not seem significant in determining labour situation.

For women, the likelihood of being employed (2), is between 8 and 10% lower, and mean hourly wages are also lower in female cells (4), they earn, on average around 20% less per hour. Circumstances determine employment rates in a 28% and hourly wages in 36%, these R-squared increase when including previous opportunity advantages.

It is important to notice that educational opportunities slightly increase the chances of being employed (3) and labour situation opportunities have also a small and positive impact on average hourly wage (5). Therefore, we can conclude that the opportunity advantages and disadvantages in education, caused by characteristics that highly determine the individual's level of studies, are somewhat carried out to affect the next labour outcomes, labour situation and hourly wage. These previous opportunities have an explanatory power around 5% for job status and almost 10% for hour salaries.

Some extra interesting things can be learned from these regressions. Some regions provide clearly better educational opportunities than Andalusia; even more outperform the reference territory in employment rates, but only a few offer higher salaries. Catalonia, Madrid, Navarre and the Basque Countries are the only regions that offer hourly wages above those of all the other autonomous communities (4), which are not significantly different from those in Andalusia, with the exception of Murcia that provides lower salaries. These results are only significant at 90% for the Basque Country, Madrid and Navarre when controlling for previous opportunity bias.

Table 4. The impact of circumstances on the outcomes

	Freq. superior studies	Freq. employed		Mean hourly wage (in logs)	
Index	(1)	(2)	(3)	(4)	(5)
Index – studies			0.435***		-0.00632**
			(0.161)		(0.00320)
Index – labour situation					0.00501***
					(0.00160)
Index – hourly wage					
Gender *(base: Men)*					
Women	7.463***	-7.852***	-9.648***	-0.249***	-0.177***
	(1.648)	(2.297)	(2.537)	(0.0424)	(0.0504)
Region *(base: Andalusia)*					
Aragón	10.03**	18.91**	14.42	-0.0352	-0.135
	(4.828)	(9.007)	(9.088)	(0.137)	(0.114)
Asturias	14.49***	12.09**	8.402	0.0495	0.0544
	(4.172)	(5.189)	(5.678)	(0.0891)	(0.0948)
Canary Islands	3.489	6.696	5.500	-0.0515	-0.0970
	(2.655)	(5.515)	(5.633)	(0.113)	(0.110)
Cantabria	16.74***	14.42	8.510	0.177	0.145
	(5.263)	(8.812)	(8.689)	(0.179)	(0.189)
Castilla León	8.588**	23.63***	20.58***	0.0912	0.0225
	(3.773)	(4.727)	(5.392)	(0.113)	(0.103)
Castilla Mancha	2.315	13.14**	11.77*	-0.0415	-0.138
	(3.103)	(6.359)	(6.345)	(0.0997)	(0.104)

	(1)	(2)	(3)	(4)	(5)
Catalonia	12.27***	21.36***	17.20***	0.203**	0.145
	(2.606)	(4.565)	(5.079)	(0.0800)	(0.0911)
Extremadura	3.409	6.901	4.072	-0.130	-0.113
	(4.563)	(4.616)	(4.427)	(0.126)	(0.135)
Galicia	9.405**	20.49***	16.03**	-0.0102	-0.0926
	(3.860)	(7.397)	(7.784)	(0.0941)	(0.0969)
Balearic Islands	2.120	24.20***	22.94***	0.172*	0.0640
	(4.614)	(4.723)	(4.518)	(0.0938)	(0.0955)
Rioja	5.913	20.78***	17.87**	-0.0562	-0.128
	(3.665)	(6.894)	(6.997)	(0.176)	(0.170)
Madrid	17.54***	25.10***	19.44***	0.247***	0.189*
	(2.535)	(3.996)	(4.640)	(0.0920)	(0.107)
Murcia	-5.391	14.07***	15.45***	-0.0631	-0.159*
	(3.446)	(4.686)	(5.136)	(0.0765)	(0.0847)
Navarre	8.021	21.59***	18.80***	0.266***	0.159*
	(5.457)	(5.980)	(6.276)	(0.0793)	(0.0874)
Basque Country	18.85***	22.52***	16.20***	0.256**	0.250*
	(3.584)	(3.799)	(4.413)	(0.124)	(0.129)
C. Valenciana	1.794	22.82***	22.02***	0.0661	-0.0554
	(3.710)	(4.091)	(4.847)	(0.0791)	(0.0928)
Family background *(base: Very bad)*					
Bad	2.285	-4.118	-4.011	-0.0372	-0.00464
	(3.244)	(5.361)	(5.201)	(0.0818)	(0.0858)
Moderately Bad	10.54***	-0.123	-1.941	0.0449	0.104
	(3.051)	(4.564)	(4.525)	(0.0785)	(0.0847)
Moderately Good	23.85***	6.698	1.844	0.153**	0.227***
	(2.870)	(4.401)	(4.518)	(0.0749)	(0.0834)
Good	38.65***	11.12**	1.622	0.276***	0.378***
	(2.841)	(4.481)	(5.320)	(0.0765)	(0.0988)
Very Good	52.61***	6.960	-10.17	0.218*	0.426***
	(3.963)	(5.748)	(9.120)	(0.112)	(0.139)
Constant	3.481	62.61***	61.31***	2.231***	2.103***
	(3.108)	(5.232)	(5.311)	(0.0987)	(0.117)
Observations	202	202	202	200	200
R-squared	0.780	0.282	0.335	0.359	0.443

Notes: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.
In columns (1), (2) and (4) the independent variables are the circumstances and the dependent variables are the outcomes per cell. In columns (3) and (5) in addition to circumstances, *previous* indexes are also included.

To sum up, we find a larger role of circumstances in determining the opportunities of education. The impact is lower for hourly wage, and even smaller for job status. This related to previous literature devoted to analyse the channels of inequality of opportunity. Palomino *et al.* (2019) find that, in 2011, a relevant share of inequality of opportunity in household income was channelled through the attained level of studies (around 16.2% of IO in Spain), whereas once controlling for education, the occupational channel is reduced (accounting for 4.3% in Spain).

The most relevant circumstance in shaping the level of studies an individual attains is the familiar economic situation when the individual was young. This variable also plays an important role in wage opportunities, but hourly salary is also highly influenced by gender, and some regions, Navarre and the Basque country, clearly provide better opportunities than the Spanish average.

Family background seems to not be significant in determining the job status[12]. Gender has some influence, but we find that the type of contract and the duration of unemployment spells are highly determined by the autonomous community. Consistently with previous literature we

[12] In further steps of this research, we plan to further analyse the job dimension of IO by checking the impact of circumstances on occupational levels, in addition to contract types and unemployment spells.

find that the regional employment differences in Spain are large, and also in the provided labour opportunities (Gorjón, de la Rica and Villar, 2020; Herrero, Villar and Guillén, 2018).

Being a woman increases educational opportunities, an advantage that is not translated neither in employment ratios nor in higher hourly wages.

4. The relevance of opportunity bias on total inequality

Once the drivers of opportunity bias and their importance and impact onto several relevant outcomes for the labour market performance have been analysed, this part of the study is devoted to quantify the extent to which total inequality is generated by inequality of opportunity (IO).

Several studies have compared the share of opportunity biases between different countries and concluded that Spain is one of the European states with the highest share of IO (Palomino *et al.*, 2019; Marrero and Rodríguez, 2012; Rodríguez, 2008).

In addition, other research shows that IO is harmful for economic growth. Economic theory suggests that overall inequality has both positive and negative effects to growth, in this setting, Marrero and Rodríguez (2013, 2019) show that IO has only negative effects. A hypothesis that is also tested by Bradbury and Triest (2016). Using US data, they find that local areas with higher intergenerational mobility display faster economic growth over the 2000–2013 and 2007–2013 periods. And also, that higher growth fosters opportunity advantages.

We want to analyse the relevance of IO on total inequality. We measure it in terms of income inequality and use yearly equivalent household income as the reference variable. We have focused the analysis of the drivers of IO on the year 2019, therefore, we first present an estimation of its importance for the same year. After that, we include an extra section where we compare the evolution of IO in two moments of time, 2011 and 2019. We compare results between these dates because it is the years when the SILC dataset includes the specific module on intergenerational transmission, allowing us to estimate IO.

The distribution of the dependent variable, yearly household income, in 2011 and 2019 is plotted in *Figure 8*. We can observe that the top decile holds more than 20% of total income in both years. These households obtain more than 30,432€ a year. Actually, the allocation of household income has not changed much during the analysed period. Half of total income is accumulated by the top three deciles. The first decile has only around 2% of total income, less than 6,186€ a year. Our goal is to understand how much of this unequal income distribution is due to circumstances, and if the share of IO has changed over the period.

The procedure we follow to achieve this objective is the same as the one used by Cabrera *et al.* (2021). It is similar to the methodology used in an extensive set of inequality of opportunity literature (Bourguignon *et al.*, 2007; Ferreira and Gignoux, 2011; Björklund *et al.*, 2012). Our approach is explained in the next section, and after that, we present our results and compare them to previous research.

Fig.8 Percentage of yearly household income that each decile holds in 2011 and 2019

4.1 Methodology

We adopt a parametric *ex-ante* approach, as defined by Van de gaer (1993), and we follow the estimation methodology proposed by Bourguignon *et al*. (2007) and Ferreira and Gignoux (2011). This procedure is suitable when the set of circumstances is considerably large and resulting subgroups might contain a small number of observations. This is the case of this study, as we want to fully use the potential of the information in the SILC module.

The main difference between Bourguignon *et al.* (2007) and Ferreira and Gignoux (2011) is how they treat the potential bias in the estimators generated by the fact that not all circumstances are observed[13]. The first use Monte-Carlo simulations to estimate bounds around the coefficients taking into account the potential bias, the latter interpret the coefficients as a lower bound of IO. Björklund *et al.* (2012) allow for heterogeneity capturing individual heteroskedasticity.

Another difference between methodologies is in the construction of counterfactual distributions. Several studies assign the same level of characteristics to all individuals[14] in the sample and compute absolute IO as the difference between total inequality and the level of inequality in the counterfactual distribution where there are no differences in circumstances. The methodology we implement, following Cabrera *et al*. (2021) and Marrero and Rodríguez (2012), relies on constructing a smoothed distribution using the fitted values of a reduced form regression of income on observed circumstances, as we explain in the next paragraphs.

[13] Some circumstances, and also *effort*, are unobservable, and probably correlated to the included circumstances. Then, the residuals are not orthogonal to the explanatory variables.

[14] For instance, Ferreira ang Guinoux (2011) and Bourguignon *et al.* (2007) assign the average level of circumstances to all individuals. Additionally, Bourguignon *et al.* (2007) also equalize parental education as if all parents attained the lowest level of mandatory education.

In this approach, the first step to quantify the share of IO is to estimate a model that captures the overall effect of (observed) circumstances on the outcome. To do so, we use individual data. The overall effect of circumstances on yearly equivalent household income, y_i, is estimated through the following OLS regression. This overall effect includes the direct effect, and also the indirect effect (the impact of the circumstances through "effort" and other unobservables, like luck, etc.).

$$\ln(y_i) = C_i\psi + \varepsilon_i$$

Then, as Cabrera *et al.* explain, if there was equality of opportunity, personal circumstances should not affect the distribution of outcomes. Consequently, inequality of fitted income, $\hat{y}_i = \exp\,(C_i\hat{\psi})$, should be zero. We can apply any inequality index to the fitted income distribution $I(\hat{\Phi}(\hat{y}_i))$, and the difference between the value we obtain and zero will be our measure of inequality of opportunity (IO). The percentage of IO over total inequality is simply obtained by dividing the inequality in the fitted distribution over total inequality.

$$\Theta_{IO} = \frac{I(\hat{\Phi}(\hat{y}_i))}{I(\Phi(y_i))} \times 100$$

The vector of observed circumstances C_i is a sub-vector of the true vector $C_i^*$of all possible circumstances (observed and unobserved) that determine an individual's advantage. Therefore, Θ_{IO} will be a lower-bound estimate of true total inequality of opportunity (Ferreria and Guignoux, 2011). We include as many circumstances as our dataset allows, but still there are many unobserved exogenous characteristics of the individual that may affect the outcome.

For the inequality index we use Gini coefficient as it is the most commonly used to assess inequality and because we are not interested in any decomposition. In addition, Brunori *et al.* (2019) argue that Gini is the best inequality index to measure IO. Comparing it to MLD they state that the latter is highly sensitive to extreme values and reveals a low sensitivity with respect to low levels of inequality. As we measure IO from a smoothed distribution, extreme values are removed and MDL systematically results in lower estimates than the Gini index.

Despite that, we also present results for the mean logarithmic deviation (MLD, E_0) in Annex D. Mainly, we do so for comparability reasons and also as a robustness check as Ferreira and Gignoux (2011) discuss that MLD is the index with the best properties. Moreover, Ramos and Van de gaer (2020) argue that some of the properties of the indexes are lost when estimating a log-linear regression and then transforming the fitted values back into levels. That's why, in Annex D, we also present the results for several combinations of logarithmic and non-logarithmic sample and fitted distributions, following Ramos and Van de gaer (2020) recommendations, but our results appear to be robust to these variations.

We apply this method for the total population in 2019. In addition, we repeat the procedure running two restricted regressions on two cohorts and for NUTS1 regions (grouped Spanish autonomous communities)[15]. We also include a time trend analysis by comparing 2019 results to 2011. This dynamic comparison is done with 2011-SILC as it includes the module on

[15] In this last part we use NUTS1 instead of the autonomous communities per se. Some automonous regions have a small number of observations and we want to guarantee that we have enough information in all regressions.

intergenerational transmission of poverty. Other waves do not provide a complete set of exogenous circumstances and do not allow us to replicate this part of the study.

4.2 Estimating Inequality of Opportunity

Table 5 shows the results for the regressions of the natural logarithm of household income on the wider set of circumstances presented in *Table 2*. We start by introducing the same circumstances used in the first part of the study to generate the opportunity advantage index (1), and then we add extra variables to provide more detail on the family background, as this methodology allows for it. All representing information about the characteristics of the individual and her household when she was around 14 years old. The main variables included are the maximum level of parental[16] education (2), the maximum skill of parental occupation (3) and whether at least one of the parents is foreign-born[17] (4). In addition, we also add information on whether the individual lived with both parents, the number of siblings, if the mother worked and the size of the city where the individual lived, in terms of the number of inhabitants (5).

Using these variables, all exogenous circumstances and none devoted to skills, merit, effort or luck, we are able to explain 18.7% of household income variability. Although this R-squared seems small, we have to take into account that we are only estimating household income as a result of circumstances (gender, region and characteristics of the individuals when they were around 14 years old). None of the common income determinants such as level of studies, occupation, economic sector, worked months, etc. are included[18]. This result is higher than the majority of the R-squared values for IO literature for Spain and other European countries. For instance, Marrero *et al.* (2012) find an R-square of 15%, Björklund *et al.* (2012) use an extensive dataset from Sweden and find a value slightly above 5%. Ferreira and Gignoux (2011) analyse several Latin American countries and find an R-squared of 17.5% for Colombia, results are higher for other Latin American countries, as for example, Brazil, with a value around 30%, consistent with what Bourguignon *et al.* (2007) find for the same country.

Table 5. Household income on circumstances

Equivalent income (in logarithm)		(1)	(2)	(3)	(4)	(5)
Gender *(base: Men)*						
	Women	-0.0379**	-0.0375**	-0.0380**	-0.0251*	-0.0236
		(0.0159)	(0.0158)	(0.0158)	(0.0150)	(0.0150)
Region *(base: Andalusia)*						
	Aragón	0.288***	0.297***	0.286***	0.334***	0.340***
		(0.0422)	(0.0414)	(0.0408)	(0.0357)	(0.0368)
	Asturias	0.170***	0.157***	0.145**	0.131**	0.133**

[16] In order to avoid missing observations, we include parental variables instead of separated variables for father and mother characteristics, following Björklund *et al.* (2012). If we were to include separated mother and father variables, we would lose all the individuals with missing information on one of the parents, and therefore, include only those that lived with both parents, which would lead to sample selection as living with both parents has a positive impact on household income.

[17] SILC dataset presents data for parental country of birth in three categories: the first category is for those born in Spain, the second for the EU and the third includes any other country in the world. Thus, we cannot add extra disaggregation in our regressions.

[18] These variables, tough relevant for the estimation of income determinants, are excluded because the goal of this research is to quantify inequality of opportunity. In order to do so, we use a reduced form model where only exogenous circumstances are incorporated as explanatories (see the methodology section in 4.1 for more detail).

	(0.0582)	(0.0579)	(0.0579)	(0.0545)	(0.0553)
Canary Islands	0.0475	0.0365	0.0304	0.106**	0.0987**
	(0.0457)	(0.0479)	(0.0478)	(0.0418)	(0.0429)
Cantabria	0.161**	0.162**	0.149**	0.158**	0.158**
	(0.0714)	(0.0711)	(0.0721)	(0.0727)	(0.0728)
Castilla León	0.313***	0.309***	0.302***	0.303***	0.311***
	(0.0305)	(0.0306)	(0.0310)	(0.0309)	(0.0314)
Castilla Mancha	0.0647	0.0678*	0.0638	0.0921**	0.103***
	(0.0394)	(0.0392)	(0.0396)	(0.0384)	(0.0382)
Catalonia	0.415***	0.414***	0.403***	0.460***	0.455***
	(0.0307)	(0.0303)	(0.0303)	(0.0298)	(0.0311)
Extremadura	-0.0439	-0.0286	-0.0171	-0.0524	-0.0409
	(0.0358)	(0.0357)	(0.0356)	(0.0359)	(0.0353)
Galicia	0.236***	0.231***	0.216***	0.201***	0.206***
	(0.0331)	(0.0332)	(0.0335)	(0.0332)	(0.0338)
Balearic Islands	0.356***	0.349***	0.334***	0.488***	0.487***
	(0.0435)	(0.0429)	(0.0431)	(0.0391)	(0.0405)
Rioja	0.214***	0.210***	0.200***	0.244***	0.252***
	(0.0666)	(0.0672)	(0.0675)	(0.0643)	(0.0653)
Madrid	0.393***	0.369***	0.351***	0.412***	0.412***
	(0.0338)	(0.0341)	(0.0342)	(0.0332)	(0.0355)
Murcia	0.0890**	0.102***	0.0998***	0.144***	0.143***
	(0.0358)	(0.0357)	(0.0358)	(0.0348)	(0.0347)
Navarre	0.463***	0.465***	0.458***	0.485***	0.495***
	(0.0370)	(0.0376)	(0.0379)	(0.0362)	(0.0367)
Basque country	0.475***	0.454***	0.445***	0.447***	0.447***
	(0.0371)	(0.0370)	(0.0373)	(0.0366)	(0.0367)
C. Valenciana	0.163***	0.159***	0.147***	0.201***	0.208***
	(0.0337)	(0.0341)	(0.0343)	(0.0337)	(0.0338)
Family background *(base: Very bad)*					
Bad	0.0277	0.0345	0.0263	-0.0202	-0.0193
	(0.0583)	(0.0594)	(0.0584)	(0.0537)	(0.0541)
Moderately Bad	0.206***	0.198***	0.180***	0.0784*	0.0717
	(0.0505)	(0.0517)	(0.0509)	(0.0465)	(0.0463)
Moderately Good	0.357***	0.328***	0.297***	0.135***	0.122***
	(0.0482)	(0.0495)	(0.0490)	(0.0451)	(0.0451)
Good	0.469***	0.401***	0.356***	0.175***	0.160***
	(0.0495)	(0.0506)	(0.0503)	(0.0464)	(0.0471)
Very Good	0.498***	0.405***	0.345***	0.200***	0.178***
	(0.0661)	(0.0660)	(0.0656)	(0.0617)	(0.0622)
Parents' education (*base: Secondary or less*)					
Post-secondary		0.0720**	0.0411	0.0936***	0.0883***
		(0.0292)	(0.0299)	(0.0281)	(0.0276)
Superior		0.232***	0.146***	0.185***	0.184***
		(0.0244)	(0.0279)	(0.0262)	(0.0259)
Parents' occupation (*base: Did not work*)					
Low skilled			-0.0364	-0.0386	-0.0765*
			(0.0439)	(0.0403)	(0.0425)
Middle skilled			0.0633	0.0347	-0.0132
			(0.0413)	(0.0377)	(0.0436)
High skilled			0.191***	0.163***	0.116**
			(0.0461)	(0.0425)	(0.0451)
Country of birth of the parents (*base: all Spanish)*					
At least one from the UE				-0.418***	-0.414***
				(0.0445)	(0.0445)
At least one from the rest of the world				-0.561***	-0.548***
				(0.0310)	(0.0309)
Lived with both parents (*base: No*)					0.0798* (0.0409)
Number of siblings (*base: None*)					
One or two					0.00593
					(0.0190)

	(1)	(2)	(3)	(4)	(5)
Three or more					-0.0258
					(0.0249)
The mother worked					0.0241
(base: No)					(0.0163)
Number of inhabitants					
(*base: Large*)					
Medium					0.00936
					(0.0201)
Small					-0.0312
					(0.0209)
Constant	9.027***	9.027***	9.010***	9.216***	9.194***
	(0.0501)	(0.0513)	(0.0598)	(0.0543)	(0.0635)
Observations	16,982	16,807	16,781	16,766	16,700
R-squared	0.098	0.111	0.116	0.185	0.187

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

Gender has a negative impact, but it is not significantly different from zero once the other variables are introduced. This is due to the fact that the dependent variable is household income, both men and women in the same household obtain the same level of household income. Therefore, gender effects are not properly detected.

Household income is higher in most regions compared to Andalusia, the reference in the estimation. In fact, once we control for the number of inhabitants, all territories exhibit a positive effect, with the exception of Extremadura, which is not significant. Familiar economic background has relevant and significant impact, though this diminishes when we include extra characteristics of the household, as some of the effects are generated by these variables and were masked inside family background coefficients. Despite that, its effect remains quite strong. In model (5), where all variables are included, coming from a family with a very good economic situation increases household income by 17.8% compared to coming from a very bad economic background.

Parental studies are also relevant and significant. If at least one parent had superior education, current household income of the individual increases by 18.4% compared to those who's maximum parental education was secondary or less. Having parents with post-secondary studies has also a positive effect of 8.8%. If at least one parent worked in a high-skilled job this has also a positive effect on nowadays individuals' household income, which increases by 11.6% compared to those whose parents did not work.

These results are consistent with previous literature finding family background and social class of the parents (occupation and education) among the most relevant variables in estimating household income inequality of opportunity (Marrero and Rodríguez, 2012; Cabrera *et al.*, 2021).

Although these variables are relevant, in our study the highest impact goes to parental migration. Having at least one parent that was not born in Spain decreases household income for the son or daughter. The impact depends on the origin of the parents, if at least one comes from the rest of the European Union (and the other is either from Spain or also from the EU), household income is 41.4% lower. The impact is even larger for those who have at least one parent from the rest of the world, they experience a reduction of 54.8%.

Marrero and Rodríguez (2012) also find a large impact of the country of birth in Spain. Their study is done using 2005 SILC data on 26 European countries and using country of birth of the individual, instead of parental country of birth. For Spain, they find a similar negative impact of being born in the rest of the EU on household income, and even higher for individuals coming from the rest of the world, of almost 70%. The size of this coefficient is the largest among all studied countries, and the coefficient for EU citizens is also among the most negative ones, only surpassed by Italy. These values indicate a large effect of intergenerational transmission of poverty and inequality for Spanish immigrants, between 2005 and 2019[19].

Among the extra variables, only living with both parents has a positive and significant effect at 90%. The increase in current household income is around 8%. The others are not significant, but provide a better adjustment for the coefficients of the principal circumstances.

4.3 The share of IO over total inequality

Fitted values from regression (5) are used to generate a smoothed distribution of yearly household income. This distribution represents the part of income explained by the observed circumstances. First, we calculate the Gini coefficient of yearly household income in 2019, the value we obtain is 0.33[20]. Then, we apply the Gini to the smoothed distribution and obtain a value of 0.17. If there was equality of opportunity, this is, if circumstances did not matter for the level of income, any inequality index applied to the smoothed distribution would be equal to zero. As long as the value obtained is significantly different from zero, there is opportunity bias. In the case of Spain in 2019, the observed circumstances of our study account for almost 52% of total inequality, the value is significant at all conventional levels. Therefore, more than half of Spanish inequality is generated by characteristics of the individual that are out of their control.

To the extent of our knowledge, this is the largest value of estimated IO for Spain. Cabrera *et al.* (2021) use a different dataset specifically designed to estimate inequality of opportunity in 2017 and find a value of 44%. This eight-points-difference might be due to an increase of IO during these two years, but also generated by the use of different variables, for instance, we include country of birth and find it largely relevant, whereas this information is not used in their estimation.

As a robustness check, in Annex D we present estimations of IO using the MLD as the inequality index. As mentioned above, the results for MDL are significantly smaller due to the index sensibility to extreme values and the use of a smoothed distribution to estimate IO. Despite that, we obtain a value around 23%, which is also higher than previous estimates. For instance, Cabrera *et al.* find an MDL value for 2017 around 18%. Palomino *et al.* (2019), find IO shares in Spain to be around 12% both for 2005 and 2011 EU-SILC data.

Nevertheless, although it has already mentioned, we have to underline again that these are lower bound estimates of IO. Because it is impossible to observe the true and complete vector of circumstances, adding more information on exogenous characteristics of the individual would result in more precise estimates of IO.

[19] Extra results for 2011 are presented afterwards. Regressions can be found in Annex C, where the impact of parental country of birth has a similar dimension.
[20] This is exactly the same value obtained by the Spanish Statistics Institute (INE).

Table 6. Share of IO of total inequality

Gini of equivalent income	Total	Circumstances	
	0.327981 (0.003285)	0.169957 (0.001476)	**51.82%**
Observations		16,700	
Standard errors in parenthesis			

4.4 Heterogeneity across groups

We are also interested in analysing the relative importance of IO for total inequality over several groups. We run separate regressions by age and aggregated regions[21] and apply the Gini coefficient to the fitted values of those regressions. Detailed results of the regressions can be found in *Annex C*, and the values for the weight of IO in *Tables 7 and 8*.

For young individuals, under 40 years old, the impact of circumstances is higher, almost 59%, whereas for individuals between 40 and 59 this value is almost 50%. The fact that circumstances are more important for younger individuals might be due to two reasons. On the one hand, it might be that, with age, individuals are able to partly detach from their origin and be able to gain control of their outcomes. On the other, it can be that circumstances had a lower effect in the past, so that older individuals, even when young, experienced higher social mobility, but this has diminished in the recent years. This last interpretation, would be consistent with what Herrero, Villar and Soler (2018) found when studying the impact of the economic recession on opportunities. Their results suggest that young individuals have lost the largest share of opportunities. From 2007 to 2016, individuals between 16 and 30 lost 50% of their income opportunities and 40% of the employment ones.

Table 7. Share of IO of total inequality per age groups

Gini of equivalent income	Total	Circumstances	
Under 40 (25-39)	0.316747 (0.005084)	0.186269 (0.002776)	**58.81%**
Observations		5,025	
Over 40 (40-59)	0.332767 (0.004126)	0.165737 (0.001669)	**49.81%**
Observations		11,675	
Standard errors in parenthesis			

There are also significant differences between regions. In some aggregated territories the importance of circumstances on total inequality is around 40%, this is the case for the North West and the South. In the rest of territories, this value is closer to 50%, therefore around half of total inequality is generated by observed circumstances. We can also see that there is no clear relation between the size of total inequality and the relevance of circumstances. For instance, we find the lowest value for the overall Gini coefficient in the North East, but the impact of IO is around 5 points higher there than in the North West and the South, regions with higher total inequality.

Table 8. Share of IO of total inequality per regions (NUTS1)

Gini of equivalent income	Total	Circumstances	
North West	0.319316 (0.007266)	0.129991 (0.004005)	**40.71%**
Observations		1,979	
North East	0.276020	0.128755	**46.65%**

[21] We do not use autonomous communities in order to have enough observations in each OLS regression.

	(0.006564)	(0.00434)	
Observations		2,283	
Madrid	0.345737	0.165301	**47.81%**
	(0.011030)	(0.004073)	
Observations		1,606	
Center	0.293983	0.157089	**53.43%**
	(0.005763)	(0.003371)	
Observations		2,336	
East	0.311747	0.148249	**47.55%**
	(0.005449)	(0.002945)	
Observations		4,942	
South	0.327694	0.136266	**41.58%**
	(0.006369)	(0.00328)	
Observations		2,998	
Canarias	0.308381	0.14764	**47.88%**
	(0.012159)	(0.008088)	
Observations		556	

Notes: Standard errors in parenthesis.
According to NUTS1 classification North West includes Galicia, Asturias and Cantabria. North East: Basque Country, Navarre, Rioja and Aragón. Center: Castilla León, Castilla Mancha and Extremadura. East: Catalonia, C. Valenciana and Balearic Islands. South: Andalusia and Murcia.

4.5 Dynamics of the relevance of IO

To finish out the analysis, we present the change of the relevance of IO over total inequality between 2011 and 2019[22]. A table for a detailed two-year comparison can be found in Annex E, but the most important results are summarized in the next plot. In this period, total inequality decreased by one point, from 33.9 in 2011 to 32.8 in 2019. Despite this reduction in overall inequality, observed inequality of opportunity increased in this period by 1.35 points. This produces an increase of 5.6% of the share of IO over total inequality, from around 46% in 2011 to roughly 52% eight years later.

For comparability reasons we have also included the values of IO obtained by the MDL index. Marrero and Rodríguez (2012) use the exact same methodology for 2005 EU-SILC database and fins a value for the IO Spanish share equal to 13.3%. For 2011, we find a smaller overall inequality level, but a larger level and share of IO, accounting for 19.5%[23]. The level of inequality of opportunity measured by MDL is also larger in 2019, representing, as mentioned, an even higher share of 22.8%. Therefore, both the share and level of IO have steadily increased in Spain for the last fifteen years.

Figure 9 shows summarized information on the share of IO by groups, calculated using the same procedure as for *Tables 7 and 8*.[24] A time increase on the share of opportunity bias

[22] In this section, slightly change due to the fact that the dataset for 2011 does not include information on the dimension of the city where the individual lived when she was around 14 years old, in terms of number of inhabitants. Despite that, shares obtained do not differ much. For comparability purposes between the two years, estimation methods are repeated without this variable. This way, results for both years are obtained using the exact same set of observed circumstances.

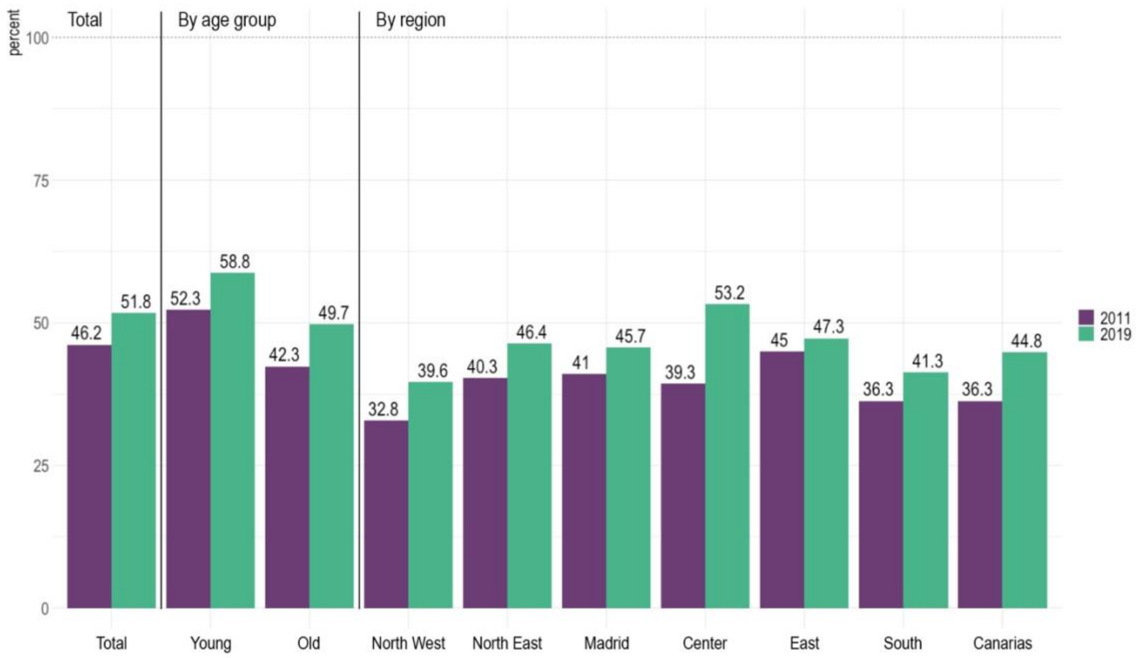
[23] We also find a larger value for 2011 IO than Palomino *et al*. (2019). They use the European version of our same dataset. The selection of circumstances is very similar, with the exception of the construction of some variables, for instance they only use father occupation and they include mother and father occupation separately. Therefore, we believe that the difference in the estimation might be due to a slight change of methodology. Whereas we estimate IO using the fitted values smoothed distribution, they construct a counterfactual and assign average circumstances to every individual and then, compute IO as the difference between total inequality and the inequality in their counterfactual distribution.

[24] Detailed results can be found in Annex E.

responsible of total inequality is experienced by every category. Whereas in 2011 IO represented from 30 to 40% of total inequality, with the exception of the young, now it is much closer to be responsible for around half total inequality.

Younger individuals burden a higher impact of their observed circumstances in both years, but their share in 2019 is the largest value of all, almost 60%. Despite that, in this period, the increase in IO is larger for individuals over 40, a rise of 7.4%. This allows us to conclude that inequality of opportunity is increasing in this period and social mobility is diminishing. It seems that, the possibility for individuals to gain control over their outcomes as they grow older is reducing, thus detachment from social origin is getting harder.

Fig.9 Share of total inequality generated by inequality of opportunity



Making a comparison between territories it stands out the percentage for the Central region which has increased by more than 14 points in eight years. The place where exogenous characteristics of the individuals matters the least is the North West region, this territory has the lowest share in both analysed years, but has increased its value quite a lot, almost 7 points. The region that has experienced the smallest change is the East, with a 2.3% rise. In addition, regional disparities have widened. In 2011, between-regions difference in the share of inequality of opportunities was around 12%, and in 2019 is closer to 14%.

To sum up, the importance of inequality of opportunity in Spain has raised during the last 15 years. We see that the shares of IO have enlarged by around six points between 2011 and 2019, the period we study. The impact of circumstances is larger for the younger population but the increase has been larger for the old. Therefore, before the relevance of circumstances decreased with age, but this gap between young and old is diminishing, indicating a larger effect of IO in the overall society. There are significant regional differences, which implies that some institutions offer higher opportunities to their citizens, whereas in other territories, individuals have more difficulties to detach their outcomes from their original background.

5. Conclusions

This study analyses the drivers of inequality of opportunity (IO) in 2019 and the importance and evolution of this kind of bias on total inequality in Spain. Opportunity disparities are relevant because they go against social justice, as they represent the part of inequality that is clearly seen as *unfair*. In addition, IO is found to be harmful for economic growth (Marrero and Rodríguez, 2013 and 2019; Bradbury and Triest, 2016). Spain is one of the European countries with larger shares of IO.

We contribute to the literature by applying an innovative methodology, the opportunity advantage index (Herrero and Villar, 2020) to analyse the impact of IO and its drivers on outcomes that are relevant for individuals' labour market performance. In addition, we measure the relevance of IO over total inequality both in 2011 and 2019. To the best of our knowledge, this is the first study measuring the share of IO in Spain using the most recently available data.

Despite the fact that total inequality has slightly reduced in Spain during the analysed period, the share of IO has increased, from around 46% in 2011 to roughly 52% eight years later. Including previous research focused on 2005 data (Marrero and Rodríguez, 2012), we can conclude that the importance of IO has steadily increased in Spain during the last fifteen years, both in absolute and relative terms.

In order to understand the drivers of such inequality, we study the effect of circumstances on the attained level of studies, labour situation and hourly wage. We find that gender, region and family socioeconomic origin explain almost 80% of the variability of educational opportunities. The impact is lower for hourly wage, almost 45%, and smaller for employment chances, around 30%. This is consistent with previous literature in finding in education an important channel of IO (Palomino *et al.*, 2019; Cabrera *et al.*, 2021).

Using the opportunity advantage index, we can quantify relative disadvantages between subgroups defined by exogenous circumstances. These circumstances describe household characteristics when the individual was young, which we link to current outcomes. We find that coming from a household with a bad economic situation, compared to a very bad economic background, reduces educational opportunities by almost 80%, and labour opportunities, both in terms of employment and salary, between 65 and 40%. The effect highly depends on the gender of the individual. Women have more educational opportunities, an advantage that is not translated neither in better labour market conditions nor in higher salaries.

Regional differences in provided opportunities are large in Spain, especially in the labour market. We find that the gap between the autonomous community with the most opportunities and the one with the least chances in labour status is around 70%. This gap is even larger for hourly wage, salary opportunities in the Basque Country and Navarre almost triple those in Extremadura.

Further research should focus on more detail on studying regional differences. Disparities can arise from different public policies that should be studied and experimented on to better understand the channels of inequality transmission and how to reduce them. In addition, some regions are attracting individuals and others see their population reduced year by year. It could be interesting to compare these internal migrations in terms of opportunities and the chances that are provided to those who move and those who stay.

Large territorial differences indicate that regional institutions can play an important role in reducing inequality of opportunity and fostering social mobility. From our results, it is necessary to intervene in an early age so that educational disparities between socioeconomic groups are diminished and not transmitted from generation to generation. Reducing opportunity bias in education will probably reduce labour market disparities, allowing a more equal access to occupational opportunities. OECD (2018) proposes investments in childcare and in family policies that balance work and family to foster social mobility, along with progressive tax systems and protection against adverse life events that result in earning losses. In addition, active labour market policies that match demand and supply are expected to reduce inequality of opportunity. Finally, we find that parental country of birth is a key variable in determining individuals' household income. Intergenerational transmission of poverty and inequality for Spanish immigrants, between 2005 and 2019, is one of the largest in Europe. Policies enhancing migrant opportunities in education and in the labour market are, thus, essential to reduce IO.

6. References

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M. & Heil, A. (2020). Why Do People Stay Poor?. STICERD - Economic Organisation and Public Policy Discussion Papers Series, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Björklund, A., Jäntti, M. and Roemer, J.E. (2012), "Equality of opportunity and the distribution of lung-run income in Sweden". *Social Choice and Welfare*, 39:675-696.

Bourguignon, F., Ferreira, F. H., and Menendez, M. (2007), "Inequality of opportunity in Brazil", *Review of Income and Wealth*, 53: 585-618.

Bradbury, K. and Triest, R. K. (2016), "Inequality of opportunity and aggregate economic performance", *The Russell Sage Foundation Journal of the Social Sciences*, 2: 178-201.

Brunori, P., Peragine, V. and Serlenga, L. (2019b), "Upward and downward bias when measuring inequality of opportunity", *Social Choice and Welfare*, 52: 635-661.

Cabrera, L., Marrero, G., Rodríguez, J.G. & Salas-Rojo, P. (2021). Inequality of Opportunity in Spain: New Insights from New Data. *Review of Public Economics* 237, 153-185.

Cantó, O. & D. Ruiz (2015), The contribution of income mobility to economic insecurity in the US and Spain during the Great Recession, *Research on Economic Inequality*, 23 : 109-152.

Choi, Á., & Gil Izquierdo, M., & Mediavilla, M. and Valbuena, J. (2018). The Evolution of Educational Inequalities in Spain: Dynamic Evidence from Repeated Cross-Sections. *Social Indicators Research*, 138.

Esping-Andersen, G., Climentada, J. & Planck, M. (2020). ¿Qué influye más en la posición social de una persona, sus habilidades o su origen familiar?. Observatori Social de La Caixa. Retrived from: https://observatoriosociallacaixa.org/es/-/posicion-social-habilidad-y-origen-familiar

Esteban, J.M., Gradín, C. and Ray, D. (2007), An extension of a measure of polarization, with an application to the income distribution of five OECD countries, *Journal of Economic Inequality* 5(1): 1-19

Ferreira, F. H. and Gignoux, J. (2011), "The measurement of inequality of opportunity: Theory and an application to Latin America", *Review of Income and Wealth*, 57: 622-657.

Gorjón, L., de la Rica, S. and Villar, A. (2020). The cost of unmeployment from a social welfare approach: the case of Spain and its regions. *Social Indicators Research*, DOI: 10.1007/s11205-020-02360-5.

Herrero, C., & Villar, A. (2018). The balanced worth: A procedure to evaluate performance in terms of ordered attributes. *Social Indicators Research*, 140(3), 1279-1300.

Herrero, C., & Villar, A. (2020). Opportunity Advantage between Income Distributions. *Journal of Economic Inequality*.

Herrero, C., Villar, A and Soler, Á. (2018), Oportunidades de Empleo y Renta en España 2007-2016. El impacto de la crisis, Ivie-Fundación Ramón Areces.

Ivie (2018),"Calculating the balanced worth vector". https://web2011.ivie.es/balanced-worth/

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

Marrero, G. A. and Rodríguez, J. G. (2013), "Inequality of opportunity and growth", *Journal of Development Economics*, 104: 107-122.

Marrero, G. A. and Rodríguez, J. G. (2019), "Inequality and growth: The cholesterol hypothesis", *ECINEQ WP*, 501.

Marrero, G. A., Rodríguez, J. G. and Van Der Weide, R. (2016), "Unequal Opportunity, Unequal Growth", *The World Bank Working Papers Series*, WP no. 7853.

OECD (2018), *A Broken Social Elevator? How to Promote Social Mobility.* OECD Publishing, Paris. Retrived from: https://doi.org/10.1787/9789264301085-en

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

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

Ramos, X. and Van de gaer, D. (2020), "Measurements of Inequality of opportunity based on counterfactuals". *Social Choice and Welfare,* 55:595-627

Rodríguez, J. G. (2008). Partial equality-of-opportunity orderings. *Social Choice and Welfare*, 31(3):435-456.

Roemer, J. E. (1998), *Equality of Opportunity*, Cambridge: Massachusetts Harvard University Press.

Roemer, J. E. (1993), "A pragmatic approach to responsibility for the egalitarian planner", Philosophy Public Affairs, 20: 146-166.

Van de gaer, D. (1993), "Equality of opportunity and investment in human capital", PhD thesis, Leuven: Katholieke University

Annex

Annex A – Robustness check comparing region of residence and region of birth using data from the Spanish Labour Force Survey in 2019.

We estimate the opportunity advantage index for level of studies and labour situation. We present the values for the index for both region variables and the difference on the evaluation computed as the index for the birth region minus the value for the region of residence. In this plot, a positive value indicates that the region offers more opportunities than the ones we can capture. On the other hand, if the value is negative, the region offers less opportunities. For example, a negative value on the difference in the level of studies index implies that citizens born in said region attain higher level of studies, but afterwards, individuals with higher levels of studies move to other regions, or the region attracts individuals with lower educational levels.

Fig.A1 Opportunity bias in studies: comparison between region of residence and region of birth

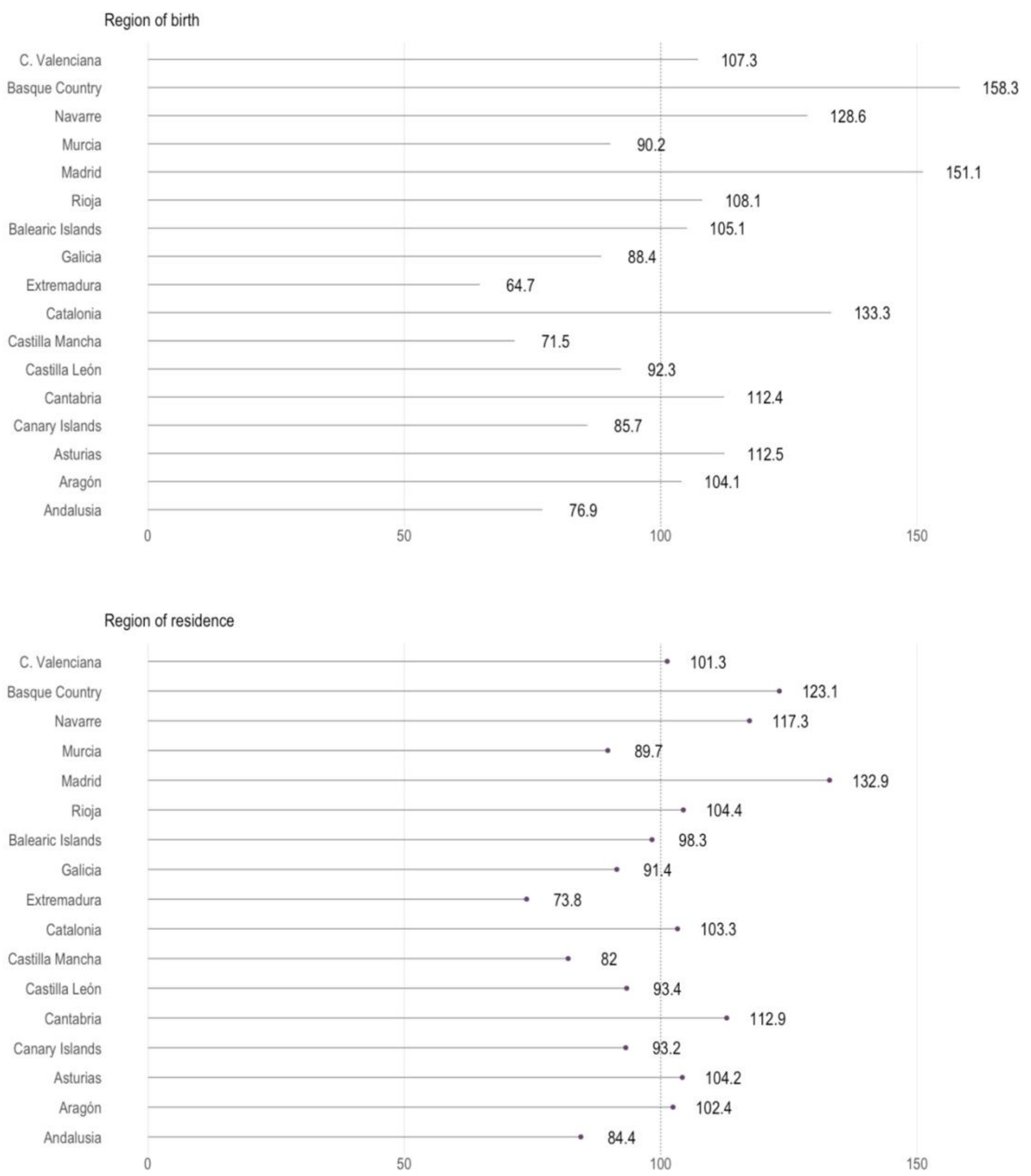
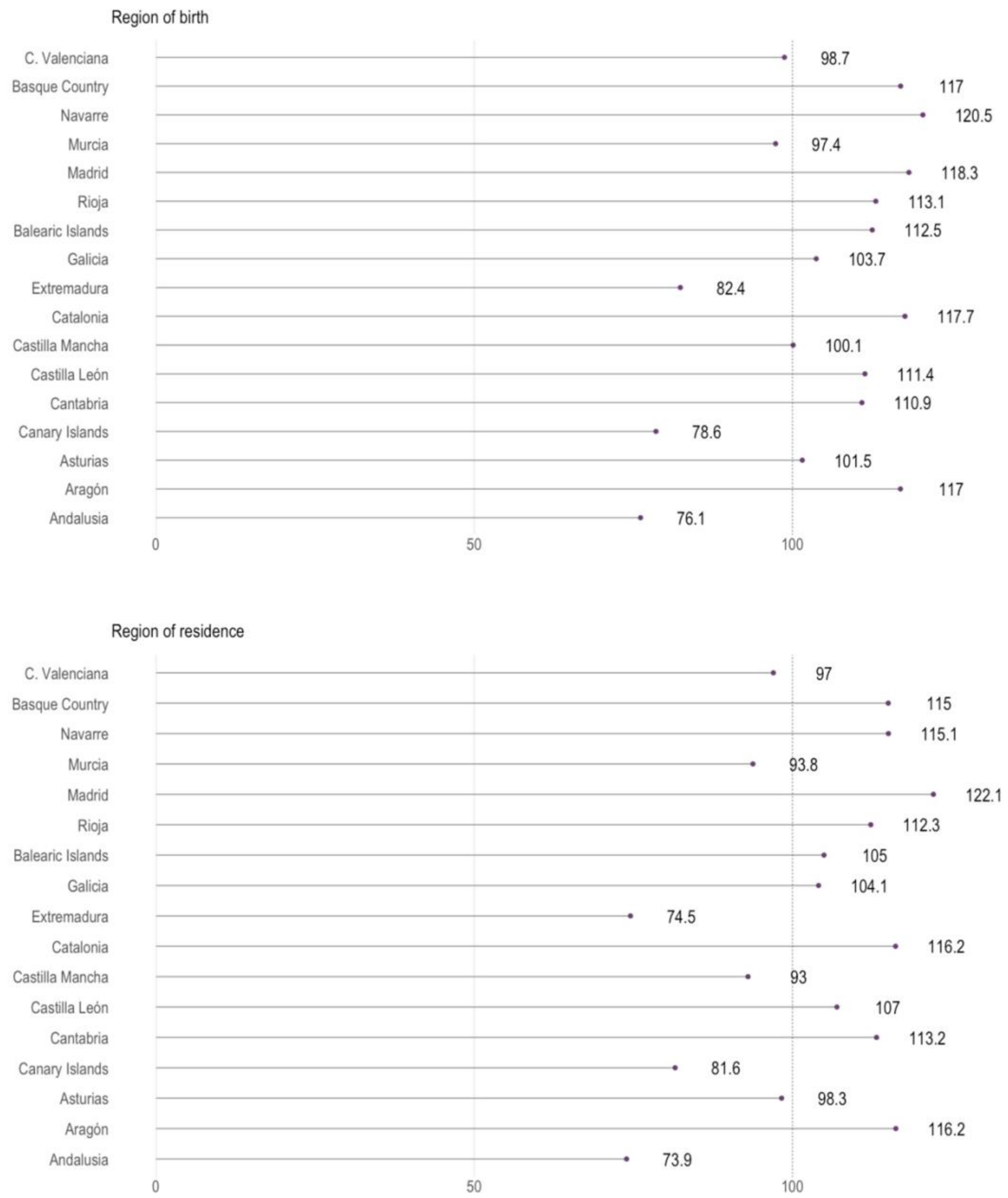
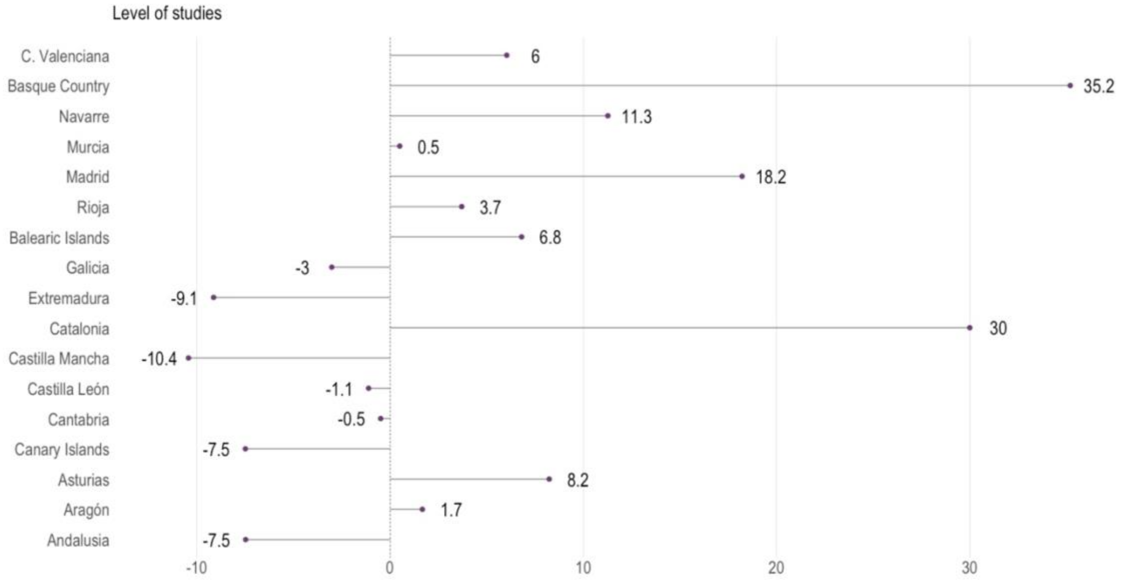


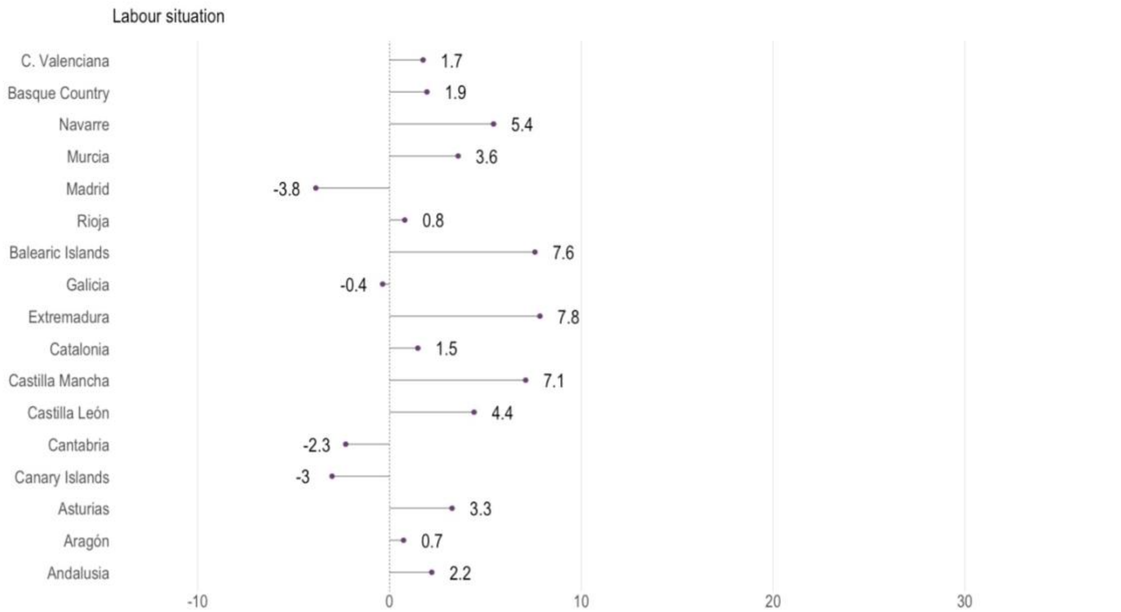
Fig.A2 Opportunity bias in labour situation: comparison between region of residence and region of birth



We can see that, in general terms, birth regional differences are slightly higher, implying that moving contributes to smoothing the territorial opportunities distribution. Nevertheless, the ranking is highly maintained. The opportunity advantage is slightly biased for level of studies when we use region of residence, but the bias is small and does not affect the ranking. Regions offering advantages above the Spanish mean and disadvantages under the mean are maintained. These differences are largely reduced for labour status.

Fig.A3 Difference in opportunity bias between region of birth and region of residence





Annex B – Plots for crossing circumstances: gender and family background.

Fig.B1 Outcome distribution by gender and family background (in percentages)

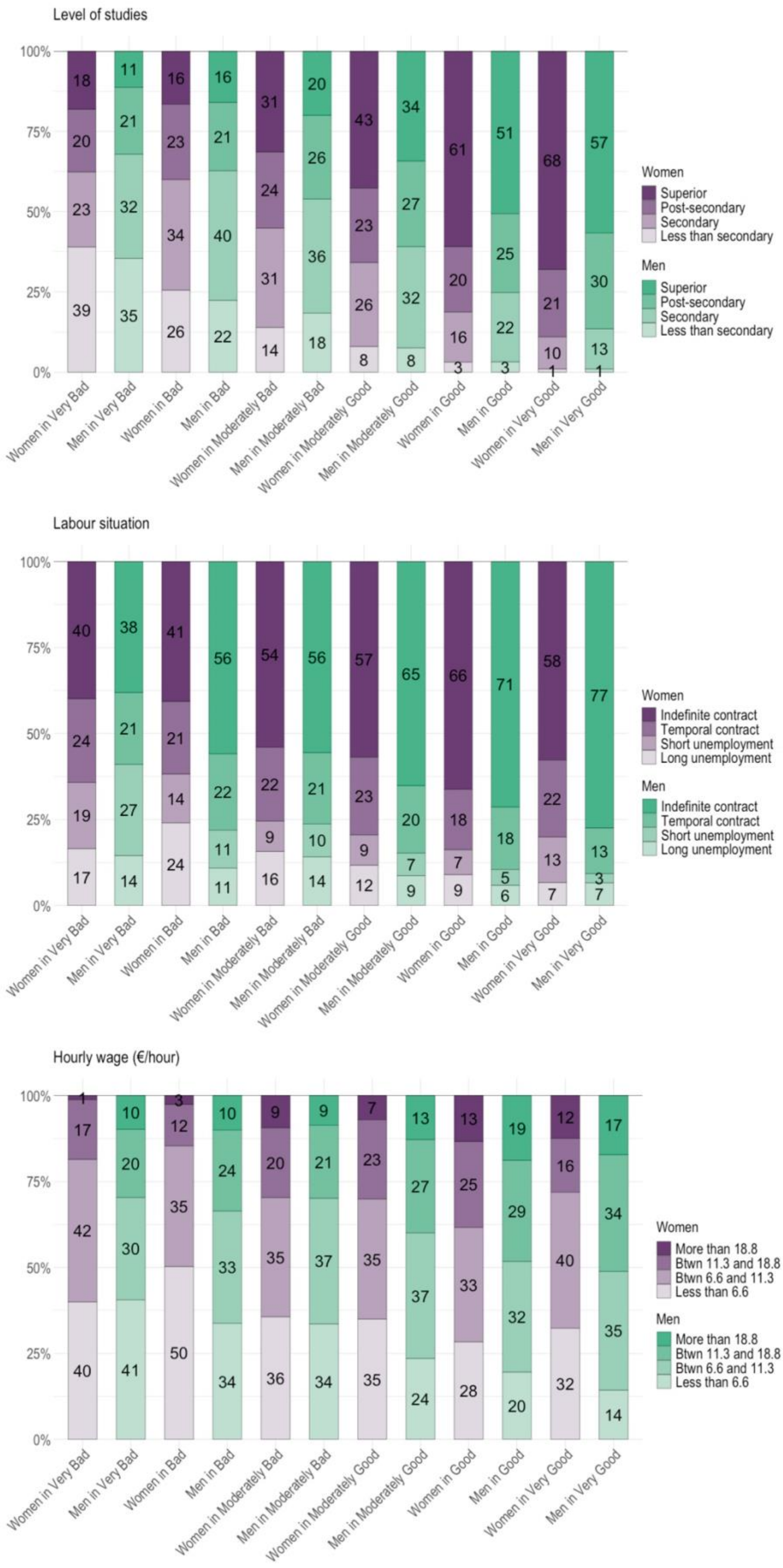
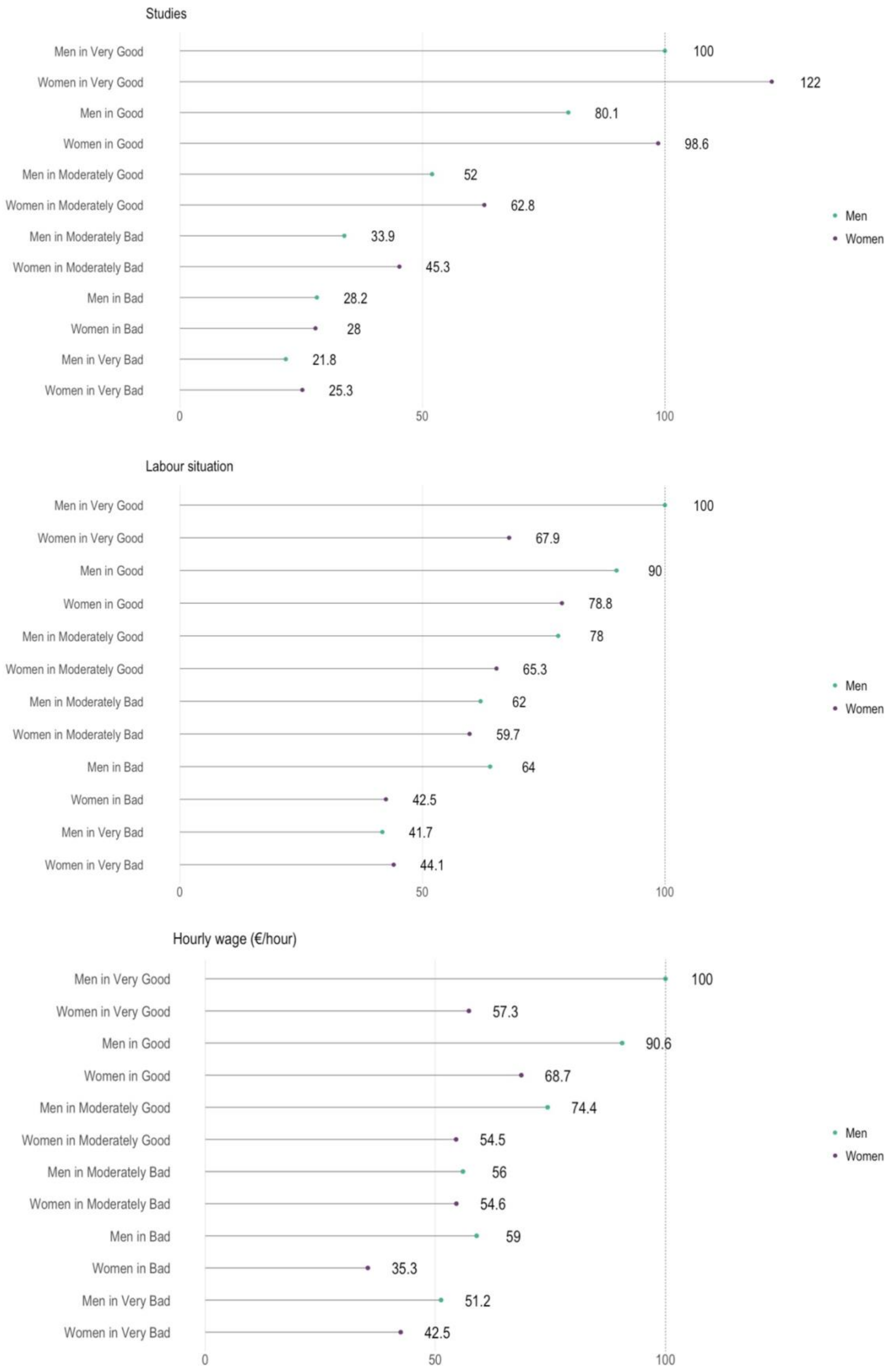


Fig.B2 Opportunity bias by gender and family background



Annex C – Additional regressions.

Household income on circumstances by age and region groups – 2019									
	Age		Region						
Equivalent income (in logarithm)	Under 40	Over 40	NW	NE	Madrid	Center	East	South	Canarias
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gender *(base: Men)*									
Women	-0.0571**	-0.00597	-0.0101	-0.0309	-0.0905**	-0.0105	-0.00928	-0.0224	0.0417
	(0.0273)	(0.0179)	(0.0415)	(0.0317)	(0.0449)	(0.0298)	(0.0268)	(0.0367)	(0.0593)
Region									
Andalusia	*base*	*base*						*base*	
Aragón	0.449***	0.278***		*base*					
	(0.0680)	(0.0441)							
Asturias	-0.00223	0.185***	*base*						
	(0.126)	(0.0518)							
Canary Islands	0.0852	0.0933*							*omitted*
	(0.0822)	(0.0488)							
Cantabria	0.177	0.136	0.0190						
	(0.142)	(0.0831)	(0.0835)						
Castilla León	0.427***	0.245***				*base*			
	(0.0562)	(0.0374)							
Castilla Mancha	0.158**	0.0680				-0.200***			
	(0.0649)	(0.0466)				(0.0356)			
Catalonia	0.455***	0.445***					*base*		
	(0.0575)	(0.0365)							
Extremadura	-0.0239	-0.0427				-0.344***			
	(0.0572)	(0.0441)				(0.0381)			
Galicia	0.227***	0.187***	0.0419						
	(0.0638)	(0.0398)	(0.0529)						
Balearic Islands	0.546***	0.444***					0.0148		
	(0.0814)	(0.0455)					(0.0363)		
Rioja	0.154	0.293***		-0.0843					
	(0.128)	(0.0732)		(0.0644)					
Madrid	0.446***	0.388***			*omitted*				
	(0.0656)	(0.0421)							
Murcia	0.209***	0.101**						0.153***	
	(0.0587)	(0.0431)						(0.0366)	
Navarre	0.512***	0.478***		0.168***					
	(0.0731)	(0.0399)		(0.0373)					
Basque country	0.504***	0.410***		0.111***					
	(0.0673)	(0.0440)		(0.0387)					
C. Valenciana	0.278***	0.159***					-0.247***		
	(0.0587)	(0.0407)					(0.0294)		
Family background *(base: Very bad)*									
Bad	0.139*	-0.0956	0.651***	-0.125	-0.0960	0.175	-0.0970	-0.0221	-0.207
	(0.0733)	(0.0726)	(0.174)	(0.104)	(0.138)	(0.170)	(0.107)	(0.100)	(0.187)
Moderately Bad	0.190***	0.00964	0.645***	0.00767	-0.0572	0.143	0.00442	0.166**	-0.246
	(0.0711)	(0.0582)	(0.168)	(0.0785)	(0.132)	(0.165)	(0.0978)	(0.0740)	(0.204)
Moderately Good	0.274***	0.0496	0.647***	-0.00751	0.0535	0.162	0.0800	0.201***	-0.0105
	(0.0666)	(0.0572)	(0.172)	(0.0710)	(0.125)	(0.162)	(0.0967)	(0.0714)	(0.180)
Good	0.290***	0.102*	0.821***	-0.0448	0.0399	0.332**	0.0949	0.186**	0.0930
	(0.0716)	(0.0586)	(0.172)	(0.0717)	(0.127)	(0.163)	(0.0994)	(0.0778)	(0.180)
Very Good	0.330***	0.105	0.854***	-0.0209	0.0208	0.250	0.212*	0.135	0.124
	(0.0968)	(0.0761)	(0.215)	(0.114)	(0.140)	(0.202)	(0.128)	(0.129)	(0.211)
Parents' education *(base: Secondary or less)*									
Post-secondary	0.0364	0.159***	0.124**	0.0536	0.0886	0.189***	0.0661	0.110	-0.0299
	(0.0443)	(0.0321)	(0.0624)	(0.0543)	(0.0654)	(0.0536)	(0.0441)	(0.0953)	(0.107)
Superior	0.134***	0.255***	0.278***	0.101*	0.183***	0.204***	0.121**	0.330***	0.126
	(0.0393)	(0.0339)	(0.0702)	(0.0541)	(0.0683)	(0.0471)	(0.0479)	(0.0652)	(0.105)
Parents' occupation *(base: Did not work)*									
Low skilled	-0.0304	-0.0947*	0.00275	0.0700	0.212	-0.0990	-0.118*	-0.238**	-0.304
	(0.0861)	(0.0485)	(0.112)	(0.108)	(0.185)	(0.0968)	(0.0690)	(0.0937)	(0.208)
Middle skilled	-0.0637	0.0197	0.104	0.0270	0.250	-0.111	-0.0318	-0.144	-0.193

	(0.0871)	(0.0495)	(0.101)	(0.103)	(0.180)	(0.0939)	(0.0627)	(0.105)	(0.200)
High skilled	0.116	0.104*	0.156	0.203*	0.377**	-0.0195	0.0975	0.0239	-0.210
	(0.0878)	(0.0535)	(0.122)	(0.106)	(0.180)	(0.100)	(0.0721)	(0.0985)	(0.217)
Country of birth of the parents									
(base: all Spanish)									
At least one from the UE	-0.347***	-0.476***	-0.578***	-0.305***	-0.475***	-0.440***	-0.228***	-0.351**	-0.623***
	(0.0659)	(0.0614)	(0.201)	(0.0605)	(0.0946)	(0.109)	(0.0738)	(0.140)	(0.0970)
At least one from the rest of the world	-0.538***	-0.527***	-0.692***	-0.749***	-0.708***	-0.635***	-0.505***	-0.366***	-0.338***
	(0.0525)	(0.0379)	(0.137)	(0.0851)	(0.0866)	(0.0946)	(0.0473)	(0.0641)	(0.104)
Lived with both parents	0.116*	0.0493	0.00299	0.129	0.0117	0.0789	0.00981	0.239*	0.0789
(*base: No*)	(0.0658)	(0.0528)	(0.0920)	(0.0884)	(0.0894)	(0.0698)	(0.0437)	(0.144)	(0.120)
Number of siblings									
(*base: None*)									
One or two	-0.00108	-0.00521	-0.0168	0.0635	-0.0619	-0.0653*	0.0125	0.0701	0.0575
	(0.0313)	(0.0227)	(0.0464)	(0.0436)	(0.0471)	(0.0352)	(0.0333)	(0.0544)	(0.0798)
Three or more	-0.0521	-0.0492*	0.0150	0.0243	-0.127*	-0.164***	-0.0169	0.0362	0.201**
	(0.0538)	(0.0283)	(0.0724)	(0.0600)	(0.0697)	(0.0493)	(0.0504)	(0.0581)	(0.0940)
The mother worked	0.0408	0.0365*	0.0939**	0.0862**	-1.75e-05	0.0765**	0.0452	-0.0470	-0.00905
(base: No)	(0.0286)	(0.0211)	(0.0418)	(0.0351)	(0.0518)	(0.0329)	(0.0296)	(0.0429)	(0.0668)
Number of inhabitants									
(*base: Large*)									
Medium	0.0227	0.0126	0.127**	-0.0631	0.190***	0.0155	0.00821	-0.0494	-0.174**
	(0.0339)	(0.0244)	(0.0612)	(0.0412)	(0.0592)	(0.0404)	(0.0307)	(0.0547)	(0.0687)
Small	0.0159	-0.0437*	0.0781	-0.0352	-0.0634	-0.0289	-0.0557	-0.0512	0.0328
	(0.0353)	(0.0260)	(0.0608)	(0.0401)	(0.0769)	(0.0397)	(0.0376)	(0.0557)	(0.0851)
Constant	8.995***	9.299***	8.634***	9.580***	9.607***	9.533***	9.768***	9.082***	9.634***
	(0.114)	(0.0749)	(0.173)	(0.123)	(0.193)	(0.190)	(0.107)	(0.142)	(0.254)
Observations	5,025	11,675	1,979	2,283	1,606	2,336	4,942	2,998	556
R-squared	0.230	0.173	0.143	0.174	0.193	0.209	0.194	0.106	0.195

Notes: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.
According to NUTS1 classification North West includes Galicia, Asturias and Cantabria. North East: Basque Country, Navarre, Rioja and Aragón. Center: Castilla León, Castilla Mancha and Extremadura. East: Catalonia, C. Valenciana and Balearic Islands. South: Andalusia and Murcia.

Household income on circumstances by age and region groups – 2011

Equivalent income (in logarithm)	All	Age		Region						
		Under 40	Over 40	NW	NE	Madrid	Center	East	South	Canarias
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Gender *(base: Men)*										
Women	-0.0148	-0.0317	-0.000896	-0.0337	-0.0227	0.0343	0.0164	-0.0420*	0.00640	-0.0428
	(0.0129)	(0.0204)	(0.0164)	(0.0315)	(0.0295)	(0.0389)	(0.0275)	(0.0241)	(0.0328)	(0.0586)
Region										
Andalusia	*base*	*base*	*base*						*base*	
Aragón	0.331***	0.341***	0.316***		*base*					
	(0.0312)	(0.0538)	(0.0364)							
Asturias	0.251***	0.218***	0.270***	*base*						
	(0.0309)	(0.0515)	(0.0367)							
Canary Islands	0.0236	0.0389	0.00715							*omitted*
	(0.0390)	(0.0523)	(0.0533)							
Cantabria	0.166***	0.259***	0.0940**	-0.0975**						
	(0.0363)	(0.0558)	(0.0462)	(0.0410)						
Castilla León	0.255***	0.286***	0.232***				*base*			
	(0.0280)	(0.0442)	(0.0351)							
Castilla Mancha	0.101***	0.124**	0.0891**				-0.149***			
	(0.0308)	(0.0500)	(0.0382)				(0.0320)			
Catalonia	0.365***	0.413***	0.326***					*base*		
	(0.0262)	(0.0415)	(0.0319)							
Extremadura	0.00597	0.0331	-0.0159				-0.243***			
	(0.0352)	(0.0589)	(0.0430)				(0.0377)			
Galicia	0.176***	0.183***	0.170***	-0.0598*						
	(0.0297)	(0.0475)	(0.0368)	(0.0359)						
Balearic Islands	0.264***	0.226***	0.296***					-0.105**		
	(0.0481)	(0.0833)	(0.0492)					(0.0467)		

Rioja	0.219***	0.224***	0.210***		-0.104***					
	(0.0346)	(0.0562)	(0.0425)		(0.0365)					
Madrid	0.401***	0.470***	0.346***			omitted				
	(0.0280)	(0.0457)	(0.0328)							
Murcia	0.126***	0.123**	0.124***						0.124***	
	(0.0338)	(0.0499)	(0.0452)						(0.0361)	
Navarre	0.530***	0.509***	0.550***		0.210***					
	(0.0355)	(0.0577)	(0.0434)		(0.0371)					
Basque country	0.452***	0.515***	0.403***		0.128***					
	(0.0319)	(0.0502)	(0.0399)		(0.0341)					
C. Valenciana	0.140***	0.162***	0.120***					-0.240***		
	(0.0283)	(0.0480)	(0.0326)					(0.0263)		
Family background *(base: Very bad)*										
Bad	0.0580	-0.00617	0.0905	0.106	0.0522	0.101	0.151*	-0.170	0.117	0.292
	(0.0487)	(0.0803)	(0.0608)	(0.0961)	(0.122)	(0.110)	(0.0893)	(0.142)	(0.0811)	(0.260)
Moderately Bad	0.0872*	-0.0164	0.147**	0.0810	0.140	0.146	0.181**	-0.0391	0.0266	0.477*
	(0.0462)	(0.0758)	(0.0580)	(0.0910)	(0.124)	(0.108)	(0.0830)	(0.129)	(0.0791)	(0.259)
Moderately Good	0.168***	0.0761	0.218***	0.160*	0.154	0.270***	0.244***	0.0435	0.142*	0.379
	(0.0450)	(0.0743)	(0.0563)	(0.0865)	(0.117)	(0.103)	(0.0807)	(0.126)	(0.0780)	(0.256)
Good	0.199***	0.111	0.245***	0.194**	0.164	0.256**	0.273***	0.130	0.166**	0.372
	(0.0460)	(0.0755)	(0.0579)	(0.0915)	(0.118)	(0.105)	(0.0846)	(0.127)	(0.0840)	(0.271)
Very Good	0.136**	0.0288	0.214***	0.124	0.296**	0.272*	0.183	0.0604	-0.0461	0.316
	(0.0627)	(0.0993)	(0.0772)	(0.102)	(0.129)	(0.150)	(0.164)	(0.141)	(0.155)	(0.299)
Parents' education *(base: Secondary or less)*										
Post-secondary	0.110***	0.0673*	0.186***	0.162***	0.113**	0.137**	0.154**	0.113**	0.0279	0.139
	(0.0296)	(0.0400)	(0.0403)	(0.0526)	(0.0551)	(0.0627)	(0.0713)	(0.0464)	(0.117)	(0.108)
Superior	0.250***	0.245***	0.253***	0.297***	0.208***	0.243***	0.272***	0.237***	0.332***	-0.0199
	(0.0283)	(0.0392)	(0.0417)	(0.0579)	(0.0539)	(0.0743)	(0.0603)	(0.0485)	(0.0725)	(0.259)
Parents' occupation *(base: Did not work)*										
Low skilled	0.0202	0.0328	0.0215	0.185**	-0.0216	0.119	0.0487	0.0311	-0.0849	-0.0799
	(0.0376)	(0.0578)	(0.0493)	(0.0728)	(0.0858)	(0.117)	(0.0743)	(0.0877)	(0.0770)	(0.168)
Middle skilled	0.0956***	0.0850	0.113**	0.211***	0.0773	0.136	0.215***	0.0747	0.00886	-0.0432
	(0.0361)	(0.0539)	(0.0481)	(0.0656)	(0.0816)	(0.111)	(0.0738)	(0.0849)	(0.0736)	(0.161)
High skilled	0.207***	0.174***	0.239***	0.231***	0.124	0.272**	0.306***	0.158*	0.163*	0.330*
	(0.0394)	(0.0593)	(0.0522)	(0.0745)	(0.0905)	(0.122)	(0.0812)	(0.0873)	(0.0877)	(0.192)
Country of birth of the parents *(base: all Spanish)*										
At least one from the UE	-0.412***	-0.472***	-0.333***	-0.624***	-0.327***	-0.279***	-0.494***	-0.393***	-0.720***	-0.268**
	(0.0438)	(0.0643)	(0.0527)	(0.219)	(0.0743)	(0.0908)	(0.0826)	(0.0771)	(0.136)	(0.130)
At least one from the rest of the world	-0.500***	-0.506***	-0.485***	-0.266***	-0.585***	-0.449***	-0.393***	-0.603***	-0.445***	-0.400***
	(0.0296)	(0.0418)	(0.0422)	(0.0728)	(0.0496)	(0.0661)	(0.0525)	(0.0417)	(0.0963)	(0.155)
Lived with both parents	-0.0285	0.0312	-0.0775**	-0.0434	0.0156	-0.0152	-0.0469	-0.0394	-0.00480	-0.154*
(base: No)	(0.0249)	(0.0378)	(0.0325)	(0.0476)	(0.0540)	(0.0737)	(0.0549)	(0.0508)	(0.0603)	(0.0899)
Number of siblings *(base: None)*										
One or two	-0.0141	-0.00329	-0.0233	-0.0208	-0.00636	-0.0241	-0.0338	0.0255	-0.0494	-0.0772
	(0.0148)	(0.0229)	(0.0188)	(0.0333)	(0.0333)	(0.0407)	(0.0338)	(0.0298)	(0.0378)	(0.0776)
Three or more	-0.0629***	-0.0855**	-0.0603**	-0.0663	-0.0306	-0.105	-0.0716	-0.0127	-0.0916*	-0.0624
	(0.0217)	(0.0389)	(0.0263)	(0.0524)	(0.0455)	(0.0654)	(0.0479)	(0.0440)	(0.0485)	(0.0872)
The mother worked	-0.00128	0.0443**	-0.0349	0.0114	0.0159	0.0546	0.0311	0.0101	-0.0641	-0.0979
(base: No)	(0.0149)	(0.0213)	(0.0216)	(0.0333)	(0.0337)	(0.0432)	(0.0367)	(0.0273)	(0.0390)	(0.0760)
Constant	9.192***	9.195***	9.203***	9.341***	9.520***	9.404***	9.275***	9.702***	9.304***	9.238***
	(0.0594)	(0.0923)	(0.0766)	(0.122)	(0.128)	(0.140)	(0.110)	(0.164)	(0.105)	(0.302)
Observations	15,731	5,916	9,815	2,169	2,664	1,514	2,517	3,529	2,538	800
R-squared	0.175	0.222	0.147	0.105	0.150	0.139	0.143	0.180	0.132	0.114

Notes: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.
According to NUTS1 classification North West includes Galicia, Asturias and Cantabria. North East: Basque Country, Navarre, Rioja and Aragón. Center: Castilla León, Castilla Mancha and Extremadura. East: Catalonia, C. Valenciana and Balearic Islands. South: Andalusia and Murcia.

Annex D – Robustness check for the estimates of the share of IO on total inequality

We present several procedures to estimate the share of IO. On the one hand we present estimations using MDL as the inequality measure. As expected, and mentioned in the study, the results for MDL are smaller. This is due to the fact that MDL index is sensible to extreme values and we measure inequality of opportunity by means of a smoothed distribution, which, by definition, has no extreme values (Brunori *et al.*, 2019).

On the other hand, Ramos and Van de gaer (2020) argue that some of the properties of the inequality indexes may be lost when combining logarithmic and non-logarithmic sample and fitted distributions.

The estimates used in the discussion are obtained from a log-linear regression, where household income, the dependent variable, is included in logarithms. Then, we transform the fitted values back to non-logarithm $\exp\left(\widehat{\ln(y_i)}\right)$. We apply our inequality index to the transformed fitted values and compare the result with total inequality measured directly by applying an index to y_i. The results obtained by this approach are highlighted in the tables.

In addition, we present several combinations. In the first column of each index, we compare total income, y_i, and fitted values in levels. The difference is that we can obtain the smoothed distribution from a linear-linear regression, without the need to transform the fitted values from logarithms to levels, or from a log-linear regression. We can see that the difference between these two approaches results into a maximum difference of 2.2%.

In column two of each index and year, the method first transforms household income into logarithms and applies the inequality index into the logarithm distribution. Then, this value is compared to the fitted distribution obtained from a log-linear regression, without the need to transform it back to levels. Following this approach we obtain smaller proportions of IO, but not extremely different. Smaller valuations are generated due to the fact that a logarithmic distribution reduces the presence of extreme values, diminishing the level of inequality in such distributions. The main problem with this approach is that we may be able to interpret the percentage of IO, but the inequality levels obtained from applying any inequality index to logarithmic incomes are not interpretable.

Overall, we can conclude that results are robust, and that IO has increased in the analysed period.

Share of inequality of opportunity over total inequality – 2019				
	Gini		**MDL**	
Total inequality (in levels or logarithms)				
	$\boldsymbol{y_i}$	$\mathbf{ln}(\boldsymbol{y_i})$	$\boldsymbol{y_i}$	$\mathbf{ln}(\boldsymbol{y_i})$
	0.327981 (0.003285)	**0.039238** (0.000516)	**0.207651** (0.004938)	**0.003536** (0.000357)
Smoothed distribution				
Dependent variable in levels y_i				
Fitted values	$\hat{y}_i$		$\hat{y}_i$	
	0.162626 (0.001642)		**0.048742** (0.001269)	
	49.58%		**23.47%**	
Dependent variable in logarithms $\ln(y_i)$				
Fitted values	$\exp\left(\widehat{\ln(y_i)}\right)$	$\widehat{\ln(y_i)}$	$\exp\left(\widehat{\ln(y_i)}\right)$	$\widehat{\ln(y_i)}$
	0.169957 (0.001476)	**0.018451** (0.000179)	**0.047414** (0.000856)	**0.000544** (0.000011)
	51.82%	**47.02%**	**22.83%**	**15.38%**

Share of inequality of opportunity over total inequality – 2011				
	Gini		**MDL**	
Total inequality (in levels or logarithms)				
	$\boldsymbol{y_i}$	$\mathbf{ln}(\boldsymbol{y_i})$	$\boldsymbol{y_i}$	$\mathbf{ln}(\boldsymbol{y_i})$
	0.338779 (0.003459)	**0.038381** (0.000411)	**0.203786** (0.004369)	**0.002758** (0.000089)
Smoothed distribution				
Dependent variable in levels y_i				
Fitted values	$\hat{y}_i$		$\hat{y}_i$	
	0.15367 (0.001521)		**0.041611** (0.001147)	
	45.36%		**20.42%**	
Dependent variable in logarithms $\ln(y_i)$				
Fitted values	$\exp\left(\widehat{\ln(y_i)}\right)$	$\widehat{\ln(y_i)}$	$\exp\left(\widehat{\ln(y_i)}\right)$	$\widehat{\ln(y_i)}$
	0.156425 (0.001313)	**0.016736** (0.000162)	**0.039707** (0.000726)	**0.000447** (0.000009)
	46.17%	**43.60%**	**19.48%**	**16.21%**

Annex E – Comparison table for changes in IO and subgroup heterogeneity

Gini of household equivalent income

	2011			2019		
	Total	**Circumstances**		**Total**	**Circumstances**	
	0.338779 (0.003459)	0.156425 (0.001313)	**46.17%**	0.327981 (0.003285)	0.169743 (0.001478)	**51.75%**

Standard errors in parenthesis

Gini of household equivalent income – Age groups

	2011			2019		
	Total	**Circumstances**		**Total**	**Circumstances**	
Under 40 (25-39)						
	0.332572 (0.006117)	0.173917 (0.002233)	**52.29%**	0.316747 (0.005084)	0.186164 (0.002780)	**58.77%**
Over 40 (40-59)						
	0.343419 (0.003922)	0.145414 (0.001554)	**42.34%**	0.332767 (0.004126)	0.165377 (0.001672)	**49.70%**

Standard errors in parenthesis

Gini of household equivalent income – Regions (NUTS1)

	2011			2019		
	Total	**Circumstances**		**Total**	**Circumstances**	
North West						
	0.298075 (0.007053)	0.097865 (0.003292)	**32.83%**	0.319316 (0.007266)	0.126363 0.004001)	**39.57%**
North East						
	0.31167 (0.006564)	0.125756 (0.003084)	**40.35%**	0.27602 (0.006564)	0.128065 0.004326)	**46.40%**
Madrid						
	0.329495 (0.011596)	0.135169 (0.003154)	**41.02%**	0.345737 (0.01103)	0.157932 0.004106)	**45.68%**
Center						
	0.326607 (0.007256)	0.12847 (0.002583)	**39.33%**	0.293983 (0.005763)	0.156487 0.00338)	**53.23%**
East						
	0.330747 (0.006396)	0.148732 (0.002465)	**44.97%**	0.311747 (0.005449)	0.147394 0.002959)	**47.28%**
South						
	0.348192 (0.007984)	0.126369 0.003696)	**36.29%**	0.327694 (0.006369)	0.135314 0.003286)	**41.29%**
Canarias						
	0.343239 (0.013832)	0.1246 0.005713)	**36.30%**	0.308381 (0.012159)	0.138224 0.008447)	**44.82%**

Standard errors in parenthesis
According to NUTS1 classification North West includes Galicia, Asturias and Cantabria. North East: Basque Country, Navarre, Rioja and Aragón. Center: Castilla León, Castilla Mancha and Extremadura. East: Catalonia, C. Valenciana and Balearic Islands. South: Andalusia and Murcia.